Photometric redshifts of faint X-ray sources: paving the way towards the study of AGN/galaxy co-evolution over cosmic time

Li-Ting Hsu

München 2015

Photometric redshifts of faint X-ray sources: paving the way towards the study of AGN/galaxy co-evolution over cosmic time

Li-Ting Hsu

Dissertation
an der Fakultät für Physik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Li-Ting Hsu
aus Taoyung, Taiwan

München, den 04.11.2015

Erstgutachter: Prof. Dr. Kirpal Nandra
Zweitgutachter: Prof. Dr. Jochen Weller
Tag der mündlichen Prüfung: 16.12.2015

Contents

Abstract **xv**

Abstract **xviii**

1 Introduction **1**
1.1 Basic properties of active galactic nuclei 2
1.1.1 AGN unified model 2
1.1.2 Properties of unobscured/type 1 AGNs spectra 2
1.1.3 Properties of obscured/type 2 AGNs spectra 6
1.2 AGN evolution 8
1.2.1 BH growth in the local Universe 8
1.2.2 AGN evolution and BH growth in the distant Universe 10
1.3 Co-evolution of AGN and their hosts 14
1.4 Techniques for selecting high-z sources 16
1.5 The challenges that we are facing 19
1.5.1 SED templates for AGNs 20
1.5.2 Multi-wavelength associations 21
1.6 The outline of this work 23

2 The fields: CANDELS/GOODS-S, CDFS, and ECDFS **25**
2.1 Data set 27
2.1.1 X-ray data 27
2.1.2 Photometric data from UV to MIR 27
2.1.3 Spectroscopic Data 31
2.2 X-ray to optical/NIR/MIR Associations 33
2.2.1 Comparing X-ray Catalogs 33
2.2.2 AGN counterparts determination method 35
2.2.3 Results 37
2.2.4 Comparison to previous results in Area 1+2 40

3 Photometric redshift in the Chandra Deep Field South **47**
3.1 Photo-z for Non X-ray detected Galaxies 48
3.1.1 Input catalogs and SEDs library 48

3.1.2 Results . . . 49
3.1.3 Discussion . . . 54
3.2 Photo-z for the X-ray selected sources . . . 61
3.2.1 Template Library Methods . . . 61
3.2.2 Population-dependent SED Libraries . . . 61
3.2.3 Constructing AGN-Galaxy Hybrid Templates . . . 63
3.2.4 Resuts . . . 65
3.2.5 Discussion . . . 74

4 Color-magnitude relations for AGNs and galaxies 77
4.1 Dust correction and AGN decomposition . . . 78
4.1.1 Dust de-reddening . . . 78
4.1.2 AGN-host decomposition . . . 78
4.2 Results . . . 82
4.3 The effects of dust and AGN contribution on the host colors . . . 89
4.4 Discussion . . . 95
4.4.1 Comparison with previous works . . . 95
4.4.2 Global evolution of AGN host color . . . 96
4.5 Summary . . . 98

5 High-redshift AGN 99
5.1 Color selection techniques for galaxies and AGNs . . . 99
5.1.1 Lyman break galaxies (LBG) . . . 100
5.1.2 Bzk-like color-color selection . . . 103
5.1.3 Results . . . 103
5.2 High-z selection via *zbest* . . . 103
5.3 High-z selection via $P(z)$ . . . 105
5.4 Comparison between techniques for selecting high-z sources . . . 110
5.4.1 $\mathrm{LBG_{V09}}$ vs. $BzK/VJL/iHM$ . . . 110
5.4.2 $P(z)$ vs. $\mathrm{LBG_{V09}}$ . . . 113
5.4.3 $P(z)$ vs. $BzK/VJL/iHM$ . . . 113
5.4.4 Optimizing the efficiency of color selection techniques defining high-z sources . . . 114
5.5 High-redshift ($z > 3$) candidates selected by $P(z)$ . . . 117
5.5.1 Sample selection . . . 117
5.5.2 Comparison with previous results . . . 123
5.6 Conclusions . . . 132

6 Summary and prospects for the future 135

A Released Catalogs 139
A.1 Cross ID reference table . . . 139
A.2 X-ray source list in ECDFS . . . 141

A.3 Photometry of X-ray sources 141
A.4 Redshift catalog 142

Acknowledgements **153**

List of Figures

1.1 AGN unified model . . . 3
1.2 Panchromatic view of an AGN . . . 4
1.3 Schematic representation of the unobscured and obscured quasars. . . . 7
1.4 Typical obscured AGN X-ray spectrum . . . 7
1.5 AGN fraction as a function of BH mass . . . 9
1.6 AGN fraction as a function of stellar mass . . . 10
1.7 The space density of luminous quasars as a function of redshift . . . 11
1.8 The number density of type 1 AGNs as a function of redshift in X-ray luminosity bins . . . 12
1.9 The cosmic star formation history . . . 13
1.10 Comparison of the star formation history with the mass accretion rate on to SMBHs . . . 13
1.11 The CMD for the R-selected galaxies from Bell et al. (2004) . . . 15
1.12 Rest-frame color-magnitude diagram of AGN host galaxy colors in the ECDFS 16
1.13 Stellar mass?star formation rate relation at $1.5 < z < 2.5$ . . . 17
1.14 Color as a function of redshift for galaxies . . . 18
1.15 Scheme of quasar evolution . . . 22

2.1 Major areas defined in ECDFS. . . . 26
2.2 Broad-band filter transmission curves. . . . 28
2.3 Subaru medium-band filter transmission curves. . . . 28
2.4 H-band magnitude as a function of spec-z . . . 32
2.5 Coordinate differences between X11 and R13 X-ray catalog . . . 36
2.6 Coordinate differences between L05 and V06 X-ray catalogs . . . 37
2.7 Cumulative fraction of the posterior p for the possible counterparts . . . 38
2.8 Three examples of multiple H-band associations . . . 38
2.9 Flow chart for four cases of X-ray to optical/NIR/MIR associations. . . . 39
2.10 Multi-wavelength images of the seven X11 sources with newly found counterparts . . . 41
2.11 Multi-wavelength images of the source R-57 (=X-234) . . . 42
2.12 Cutouts of 17 unidentified X-ray sources . . . 44

3.1 Photo-z vs spec-z for normal galaxies in Area 1 . . . 51

3.2 Photo-z vs spec-z for normal galaxies (i.e., non X-ray detections) in Areas 2+3. 52
3.3 photo-z distribution for normal galaxies in Area 1+2+3 53
3.4 Photo-z vs spec-z for normal galaxies in Area 1 55
3.5 J-band magnitude vs. error 56
3.6 Distribution of $(z_{\rm p} - z_{\rm s})$ with and without IB data 57
3.7 Distribution of $(z_{\rm p} - z_{\rm s})$ for BzK-selected star-forming galaxies with and without IB data 57
3.8 $(z_{\rm p} - z_{\rm s})$ distribution in various redshift bins for normal galaxies in Area 1 58
3.9 Comparisons of the $(z_{\rm p} - z_{\rm s})$ distribution with/without the use of IBs or emission lines 59
3.10 Distributions of Δz_p for pairs of galaxies 62
3.11 Soft X-ray flux distributions 64
3.12 Galaxy templates from Bender et al. (2001) 65
3.13 Galaxy templates in use for photo-z estimation 66
3.14 Two examples of hybrid templates. 66
3.15 SEDs for all templates in Lib-EXT. 67
3.16 SEDs for all templaes in Lib-PT. 68
3.17 Comparison of photo-z to spec-z with and without $\rm TFIT_{CANDELS+IB}$ photometry. 70
3.18 Comparison of photo-z to spec-z for all X-ray sources in Areas 1+2+3. . . 71
3.19 Distribution of 0.5–8 keV observed-frame X-ray luminosity as a function of redshift 72
3.20 Spectral energy distributions for typical galaxies and AGNs 75
3.21 Comparison between photo-z computed with and without UV data for the X-ray sources 76

4.1 Four extinction laws used in our SED fitting 79
4.2 Illustration of AGN-galaxy decomposition 81
4.3 Color-magnitude diagram for normal galaxies and AGN hosts 84
4.4 Color distributions for the normal galaxies 85
4.5 Color distributions for the AGNs 85
4.6 Comparison of color distributions between galaxies and AGNs with the corrections 87
4.7 Comparison of color distributions between galaxies and AGNs with the corrections ($M_V < 20$) 87
4.8 Comparison of color distributions between galaxies and AGNs without the corrections 88
4.9 Comparison of color distributions between galaxies and AGNs without the corrections ($M_V < 20$) 88
4.10 Mean values of $\Delta(U - V)$ as a function of redshift. 90
4.11 Color distributions of X-ray selected sources in terms of the distances to the color cut 92
4.10 (Continued) 93

4.11 $\Delta(U-V)$ as a function of ΔM_V 94
4.12 Color distributions for the K-selected galaxies in Williams et al. (2009) . . 97
4.13 Color distributions for the normal galaxies ($H < 22.5$) 97

5.1 $\mathrm{LBG_{V09}}$-selected AGNs using the B_{450}-band dropout technique in the GOODS-S 102
5.2 BzK-selected AGNs in GOODS-S (Area 1) 104
5.3 VJL-selected AGNs in GOODS-S (Area 1) 104
5.4 iHM-selected AGNs in GOODS-S (Area 1) 105
5.5 Example of measuring the value of weight w from the $P(z)$ 106
5.6 $(C + P_s)$ as a function of w_{th} in redshift bins 109
5.7 Selection efficiencies of color methods as a function of type of sources and redshift ranges 116
5.8 Examples of SED and $P(z)$ of high-redshift candidates 118
5.9 Comparison of our photo-z results obtained using different magnitude priors 124
5.10 Multi-wavelength images of the X-ray source (X-645) 125
5.11 Extremely high-z($z > 5$)candidates from Fiore et al. (2012) 125
5.10 (Continued) 126
5.9 (Continued) 127
5.8 (Continued) 128

A.1 Two examples of SED fitting for a normal galaxy and a X-ray-detected AGN 140

List of Tables

1.1 Color selections of high-z sources 18

2.1 Photometric Data in $\mathrm{TFIT_{CANDELS+IB}}$ catalog 29
2.2 Photometric Data in MUSYC+TENIS catalog 30
2.3 Catalogs used for redshift estimation and counterpart identification. 34
2.4 Results of X-ray to optical/NIR/MIR associations in ECDFS 42

3.1 Photo-z Quality for Normal Galaxies in Area 1+2+3 50
3.2 Comparison of Photo-z Results for Normal Galaxies in Area 1 50
3.3 Influence of IB and EM on photo-z quality 60
3.4 Photo-z Scatter from Pair Statistics 60
3.5 The list of AGN-galaxy hybrids in Lib-EXT and Lib-PT. 69
3.6 Photo-z Quality for X-ray Sources 73

4.1 Mean values of $\Delta(U-V)$ derived from the correction of dust and AGN contribution for AGNs and galaxies. 90

5.1 Information from the SED fitting of $\mathrm{LBG_{V09}}$-selected sample in Area 1 . . 101
5.2 The values of w_{th}, completeness (C) and purity (P_s) for which ($C+P_s$) is maximum for X-ray and non-X-ray detected samples in Area 1. 108
5.3 Completeness (C) and purity (P_s) for AGN and galaxy samples selected by $P(z)$, *zbest* and $\mathrm{LBG_{V09}}$ $/BzK/VJL/iHM$ techniques in Area 1 111
5.4 Comparisons between the original works and our works on the filters, depth and regions used for color techniques. 112
5.5 Properties of $P(z)$-selected sample that are not included in the $\mathrm{LBG_{V09}}$ sample in Area 1. 113
5.6 Redshift range optimized by maximizing ($C+P_s$) for AGN and galaxy samples selected by $P(z)$, *zbest* or $\mathrm{LBG_{V09}}/BzK/VJL/iHM$ techniques in Area 1 115
5.7 Number of high-redshift ($z \geq 3$) X-ray sources 117
5.8 High-redshift source list with $zbest > 3$ 119
5.8 (Continued) 120
5.9 High-redshift source list with $zbest < 3$ and $(z_{\mathrm{p}}+3\sigma) \geq 3$ 121
5.9 (Continued) 122

5.10 Non-X-ray-detected high-z sources selected by Fiore et al. (2012). 129
5.11 X-ray detected high-z sources selected by Fiore et al. (2012). 130
5.12 High-z ($z > 5$) candidates from Fiore et al. (2012) 131
5.13 Newly-found high-z ($z > 3$) X-ray sources in CANDELS/GOODS-S from Hsu et al. (2014) 131

A.1 Column description of the cross ID reference catalog 140
A.2 X-ray source list 141
A.3 Photometry of X-ray sources 142
A.4 Redshift catalogs 143

Zusammenfassung

Aktive Galaktische Kerne (active galactic nuclei, hier: AGN) sind super-massive Schwarze Löcher (supermassive black holes, hier: SMBH) mit aktiver Akkretion im Zentrum massiver Galaxien. Die engen Korrelationen von $M_{BH} - \sigma$ und $M_{BH} - M_{bulge}$ stellen den Einfluss von SMBH auf Galaxien dar (z.B. Magorrian et al., 1998; Gebhardt et al., 2000; Marconi & Hunt, 2003). Viele vorherige Studien zu Farb-Magnituden-Relationen zeigen auf, dass sich die Farben von AGN-Muttergalaxien im Übergang von der blauen Wolke zur roten Sequenz befinden (z.B. Nandra et al., 2007; Hickox et al., 2009) und interpretieren dieses Ergebnis als Beweis für AGN-Feedback (z.B. Silk & Rees, 1998; King, 2003). Akkretionsprozesse bauen also nicht nur schwarze Löcher auf, sondern haben - indem sie die Sternenbildung anstoßen oder unterdrücken - auch einen starken Einfluss auf die Umgebung.

Diese Korrelationen zeigen die Notwendigkeit zur Untersuchung der Koevolution von AGN und Muttergalaxie auf. Dies kann z.B. durch die Studien des Aktivitäts-Zyklus eines AGN geschehen (hier: duty cycle, d.h. die Zeitspanne während das SMBH aktiv ist), welcher anhand der Verteilung von AGN Populationen innerhalb aller Galaxien in Abhängigkeit von der kosmischen Zeit abgeschätzt werden kann. Demographische Untersuchungen dieser Populationen erfordern gut erprobte Auswahlkriterien und akkurat bestimmte Rotverschiebungen. Bei hoher Rotverschiebung ist die Anzahl von Quellen für umfangreiche Datensätze sehr gering und meist zu lichtschwach. Aus diesen Umständen resultieren große statistische Fehler. Aird et al. (2010) argumentieren zum Beispiel, dass die luminositätsabhängige Dichteevolution, repräsentiert durch eine abfallende Steigung am lichtarmen Ende der Röntgen-Luminositäts-Funktion (XLF) für $z \geq 1.2$, aus den extremen photometrischen Rotverschiebungsfehlern kommt. Diese Fehler resultieren aus beobachtungsbedingten Einschränkungen und unpassenden Vorlagen in der spektralen Energieverteilung (hier: templates) für die Berechnung der photometrischen Rotverschiebungen. Spektroskopische Beobachtungen sind seh zeitintensiv und für große Datensätze nur aufwändig durchführbar. Daher treten photometrische Rotverschiebungen an ihre Stelle. In deren Bestimmung fließen umfangreiche Informationen zu den AGN ein: z.B. Magnituden-Prior, hybride AGN-Galaxie Templates für den Fit spektraler Energieverteilungen und korrekte Counterparts in verschiedenen Wellenlängen. Seit einigen Jahren sind tief- und hochaufgelöste Beobachtungen in vielen Wellenlängen verfügbar. Diese ermöglichen das Erreichen höherer Rotverschiebungen und genauere Analysen der AGN Eigenschaften.

Die vorliegende Arbeit umfasst eine Studie im ECDFS-Gebiet, welches die GOODS-

S- und CDFS-Regionen umfasst. Dieser betrachtete Himmelsausschnitt enthält die tiefsten und vollständigsten photometrischen Informationen von Röntgen- bis zu Radiowellen; ebenso von Subaru und mittel-/nah-infrarot Photometrie vom Hubble Space Telescope (*HST*). Zur Berechnung akkurater photometrischer Rotverschiebungen kombinieren wir verschiedene Kataloge von UV- bis zu Infrarotwellen, nachdem diese einer astrometrischen Kalibration und einer Korrektur für unterschiedliche Fluss-Bestimmungen (z.B., total Flüsse, Aperturen-Flüsse und PSF-Photometrie) unterzogen wurden. Anschließend identifizieren wir den besten Counterpart für die Quellen aus dem 4Ms-CDFS- und 250ks-ECDFS-Katalog. Dabei beziehen wir Positionsfehler und die vielfaltigen Wellenlängenverteilungen als Prior mit ein. Wir stellen fest, dass 96% aller Röntgen-Objekte Counterparts aus anderen Wellenlängen besitzen. In einem dritten Schritt erstellen wir eine Bibliothek von hybriden AGN-Galaxie-Templates, welche für die lichtarme Röntgen-Population von CDFS geeignet ist und deren spektrale Energieverteilung bei geringer, als auch hoher Rotverschiebung simulieren kann. Für röntgenselektierte AGN erreichen wir eine Präzision der Rotverschiebung von $\sigma_{\rm NMAD} \sim 0.013$ mit einer Ausreißer-Rate von $\sim 5.3\%$. Für Galaxien ohne Röntgen Detektion liegt die Präzision der photometrischen Rotverschiebung bei ~ 0.010 mit einer Ausreißer-Rate von $\sim 4.6\%$.

Basierend auf den Resultaten der spektralen Energie-Verteilung mit unseren getesteten hybriden AGN-Galaxie Templates studieren wir die Beziehungen zwischen Muttergalaxien von AGN und normalen Galaxien. Dabei untersuchen wir das Farb-Magnituden-Diagramm im Ruhesysten (color-magnitude diagram, hier: CMD), welches die Bestimmung von stellaren Populationen unterstützt (Bell et al., 2004). Wir nehmen Korrekturen für die Farben der Galaxie und AGN-Muttergalaxien in der CANDELS/GOODS-S Region in Hinsicht auf Staub-Extinktion und/oder die Kontaminierung durch AGN vor. Dabei stellen wir fest, dass die Farben der AGN-Muttergalaxien eine gleiche Bimodalität wie die normalen Galaxien im CMD (bis zu $z \sim 2.5$) aufweisen. Die Positionen der blauen Spitzen der AGN-Datensätze sind in der kosmischen Zeit nahezu konstant. Im 4Ms-CDFS Survey begründet sich die Bimodalität der Muttergalaxien von AGN sowohl in der Korrektur für die Staub-Extinktion, als aus der Korrektur für den AGN-Beitrag. In der AGN-Population der 4Ms-CDFS überwiegen Objekte mit geringer Luminosität. Deren spektrale Energieverteilungen sind durch die Muttergalaxie dominiert und deren Farben verhalten sich wie normale Galaxien. Dennoch sind in einigen hellen Objekten die Farben der Muttergalaxien stärker durch die AGN Aktivitt als durch Staub-Extinktion beeinflusst. In diesen Galaxien ist die AGN-Korrektur im Allgemeinen zweimal größer als die Staub-Korrekturen. Daher wird die Unterscheidung zwischen AGN-Emission und der Muttergalaxien umso wichtiger in den weniger tiefen aber ausgedehnteren Röntgenbeobachtungen, wie z.B. bei XMM-COSMOS und eROSITA, da diese mehr helle AGN-Quellen enthalten.

Mit den akkuraten Rotverschiebungen für Galaxien und AGN definieren wir einen hochrotverschobenen Datensatz. Dabei verwenden wir die Informationen aus der Wahrscheinlichkeitsdichten-Funktion für Rotverschiebungen $P(z)$ anstelle der Werte für die photometrische Rotverschiebung. Wir integrieren P(z) in einem gegebenen Rotverschiebungsintervall und erhalten die Wahrscheinlichkeit für die photometrische Rotverschiebung in diesem Intervall sowie hoch rotverschobene Quellen oberhalb einer gewissen Grenze. Wir vergleichen die

"$P(z)$ -Technik" mit konventionellen Farb-Techniken, die auf die evolutionären Stadien von Galaxien ausgerichtet sind (z.B. Lyman break galaxy, BzK Farb-Farb-Selektion via Vollständigkeit und Reinheit). Wir stellen fest, dass die $P(z)$-Methode am Verlässlichsten und Effizientesten für die Selektion hochrotverschobener Quellen ist. Dies ist nicht überraschend, da dieser Datensatz die photometrischen Informationen aus der gesamten spektralen Energieverteilung miteinbezieht und photometrische Fehler berücksichtigt. Zuletzt erstellen wir einen hochrot-verschobenen Datensatz ($z > 3$) von Röntgen-Objekten in der ECDFS-Region und vergleichen diesen mit unseren vorherigen Studien. Wir erreichen bessere Röntgen-NIR-Assoziationen und berücksichten sowohl die Positionsfehler als auch die Magnituden-Verteilung. Zusätzlich erhalten wir eine akkurate photometrische Rotverschiebung durch die Verwendung hybrider AGN-Galaxie-Templates für röntgenselektierte AGN und wenden $P(z)$ für jedes Objekt an. Diese Prozeduren unterstützen die Verbesserung der Selektion von hochrotverschobenen Objekten.

Abstract

Active Galactic Nuclei (AGN) are the accreting super massive black hole (SMBH) at the center of massive galaxies. The tight $M_{BH} - \sigma$ and $M_{BH} - M_{bulge}$ correlations reveal that the host galaxies are affected by the effects of the SMBHs (e.g., Magorrian et al., 1998; Gebhardt et al., 2000; Marconi & Hunt, 2003). In addition, many works studying on the rest-frame color-magnitude relation have shown that AGN host galaxies have intermediate colors, which are considered as a transition from the blue cloud to red sequence in host-galaxy color evolution (e.g., Nandra et al., 2007; Hickox et al., 2009). Some works interpreted this result as an evidence for the AGN feedback (e.g., Silk & Rees, 1998; King, 2003), in the sense that the accretion process not only builds up the massive black holes, but also has a powerful influence on the surrounding environment, triggering or suppressing the star-forming activity in the host galaxy.

These correlations make obvious the need to investigate AGN-host coevolution. One way to understand this coevolution is to study the AGN duty cycle (i.e., the time scale that the SMBH is active), which can be obtained by estimating AGN population among all the galaxies through cosmic times. Such demographic studies require a well-sampled census and accurate redshift information. In particular at high redshift, objects are extremely faint and sample numbers are very small. This could cause large statistical errors. For example, Aird et al. (2010) argued that luminosity-dependent density evolution with a flattening faint-end slope of the X-ray luminosity function (XLF) at $z \geq 1.2$ may result from catastrophic photo-z failures caused by observational limitations and improper templates used for photo-z computation. However, spectroscopic redshifts are time-consuming and difficult to be obtained for faint sources at high redshifts. Therefore we have to rely on photometric redshifts (photo-z) techniques which need to be tuned specifically to be reliable for AGNs (i.e., proper magnitude priors, appropriate AGN-galaxy hybrid template for SED-fitting, and correct multi-wavelength counterparts). In recent years, many deep and high-resolution observations become available in multiple wavebands, specially at near/mid-infrared. This allows us to reach higher redshift, and make more accurate analysis on the multi-wavelength properties of AGNs.

In this thesis, we focus on the ECDFS area which comprise also the GOODS-S and CDFS regions. This is the portion of the sky with the deepest and most complete photometric information from X-ray to radio, including intermediate bands from the Subaru telescope, and optical/near-infrared data from the Hubble space telescope. To compute accurate photo-z using these data, first we combined multi-wavelength catalogs from UV

to infrared after the astrometric calibration and correction for the different methods of flux extraction (e.g., total fluxes, flux apertures and PSF- fitted photometry). Second we identified the best multi-wavelength counterparts for X-ray sources from the 4Ms-CDFS and 250ks-ECDFS surveys, taking into account the positional errors and multiple magnitude distributions as priors simultaneously. We found that more then 96% of X-ray sources have multi-wavelength counterparts. Thirdly we built a new library of active galactic nuclei/galaxy hybrid templates appropriate for the faint X-ray population in the CDFS to simulate the AGN spectral energy distribution from low to high redshift. For X-ray-selected AGNs, we achieved a photo-z accuracy of 0.013 with an outliers fraction of 5.3%, while for non-X-ray galaxies, the photo-z accuracy is 0.010 with an outlier fraction of 4.6%.

With the SED-fitting results of our well-trained AGN-galaxy hybrids, we further studied the galaxy and AGN host properties via the rest-frame color-magnitude diagram (CMD) which is an useful probe to trace the stellar populations (Bell et al., 2004). We made corrections for dust extinction and/or AGN contamination for the galaxy/AGN host colors in the CANDELS/GOODS-S region. We found that the AGN host colors also present bimodality in the CMD up to $z \sim 2.5$ as found in normal galaxies, and the position of the blue peaks in the AGN samples are almost constant with cosmic time. This implies a weak connection between AGN activity and star formation in the host galaxy. For the X-ray sources in the 4Ms-CDFS survey, we found that for most of the sources, the correction for dust extinction is larger than the correction for the AGN contribution. This is because the AGN population in this field is dominated by low-luminosity AGNs which have host-dominated SEDs. However for few bright sources, their host colors are strongly effected by AGN contribution rather than by the dust extinction. For these sources, the correction for AGN contribution is about two times larger than the correction for dust extinction in general. Therefore AGN/galaxy decomposition becomes more important in a shallower and wider X-ray surveys, e.g., XMM-COSMOS and eROSITA, which contains a larger fraction of bright AGNs.

Furthermore, with our accurate redshifts for galaxies and AGNs, we defined a high-redshift (high-z) sample using the redshift probability distribution function $P(z)$ rather than relying on the best-fit value of photo-z. We integrated $P(z)$ within a given redshift range to obtain the photo-z probability in that range and selected high-z sources above a given threshold. When computing the number of sources in a given redshift range, each source will not be counted as "1" but as the proportion of it. We compared this $P(z)$ technique with traditional color techniques adopted for galaxy evolutionary stages, like the Lyman break galaxy and the BzK color-color selection via sample completeness and purity. We found that the $P(z)$ technique is the most efficient and reliable method for selecting high-z sources. This is not surprising as it makes use of photometric information from the entire SED rather than using only three photometric points. Lastly, we built a high-z ($z > 3$) sources list for X-ray sources in ECDFS region, and compared our list with previous work. In our work, we made better X-ray-to-optical/NIR associations considering the positional errors and magnitude distribution. In addition, we obtained accurate photo-z using well-established AGN-galaxy hybrids for X-ray selected AGNs and applied $P(z)$ for each source. These procedures help improving on our high-z sample selection.

Chapter 1

Introduction

With the advances in telescope technology since the early 1990s, much observational evidence has accumulated that supermassive black holes (SMBH; $M_{\rm BH} > 10^6\,{\rm M}_\odot$) exist at the centers of most galaxies. Researches such as the studies of broad gravitationally redshifted Fe Kα emission line (e.g., Tanaka et al., 1995; Nandra et al., 1997; Reynolds, 1997), reverberation mapping (e.g., Krolik, 2001; Peterson et al., 2004), or gas and stellar dynamics surrounding the compact nuclei of galaxies (e.g., Ford et al., 1994; Magorrian et al., 1998; Ghez et al., 2003; Schödel et al., 2002), all point to the existence of SMBH at the centers of galaxies. Some of them are considered inactive, like the nucleus in our Milky Way, having no or marginal accretion activity. Some of them, on the contrary, have strong accretion activity which can produce powerful radiation localized in the circumnuclear region (Rees, 1984). Such energetic regions in the center of galaxies are named "Active Galactic Nuclei" (AGN).

In the entire lifetime of a SMBH, the accretion process takes place as long as accreting matter is available in the vicinity of the black holes, and thus the black holes grow (Soltan, 1982). During the accretion process, AGNs emit powerful radiation which can be observed at all wavelengths from Gamma-ray to radio, including X-ray, ultraviolet (UV), optical, infrared (IR), and sub-millimeter (sub-mm). In the following introduction, we will describe the origin of AGN emission in multiple wavebands and the typical characteristics in AGN electromagnetic spectra. In addition to the physical description of AGNs, we will give an overview of the SMBH evolution and introduce the observational evidence of the AGN and galaxy coevolution in the local and distant Universe. Moreover, we will present the factors needed to be taken into account when using the photometric and spectroscopic data to study the relation between AGN and galaxy. Lastly, we will give a brief summary of the structure of this thesis.

1.1 Basic properties of active galactic nuclei

1.1.1 AGN unified model

The general concept of the AGN unified model is that the central SMBH is surrounded by a parsec-scale accretion disk, in some cases, also accompanied by a powerful kiloparsec-scale jet. Further out from the accretion disk is the dusty torus on the scale of few tens of parsecs (pc; where 1pc = 3.26 light years). Antonucci (1993) and Urry & Padovani (1995) provided a conceptual scheme for the model as shown in the left of Figure 1.1. Based on this scenario, different types of AGN can be identified according to their spectroscopic features caused by the orientation effect of viewing the smooth-density torus. Broad line regions (BLR) and narrow line regions (NLR) can both be observed from face-on view of torus, thus broad and narrow emission lines are presented in the spectra. Such AGNs are classified as "type 1". However, with edge-on view, AGNs are classified as "type 2" that broad lines are obscured by torus and only narrow lines are presented in the spectra (see the middle spectra in Figure 1.1).

However, in the last few years, the model with smooth-density torus has been gradually substituted by a clumpy torus model which is a continuous cloud distribution from BLR toward the outer plane (Nenkova et al., 2002, 2008a,b). In this scenario, the spectroscopical classification remained but AGN types are defined by the probability for directly viewing the AGN rather than due to the orientation effect (Figure 1.1). Following, we describe the main continuum properties of type 1 and type 2 AGNs over the entire electromagnetic spectra, which are relevant to the need for computing photo-z in the later Chapter 2 and 3.

1.1.2 Properties of unobscured/type 1 AGNs spectra

Elvis et al. (1994) built a mean spectral energy distribution (SED) by collecting the multi-wavelength observations of luminous AGNs from hard X-ray to radio. This SED reveals distinguishing spectral features of unobscured/type 1 AGN, such as "big blue bump" , "IR excess", strong and broad emission lines, and "Compton hump" (Figure 1.2). These phenomena provide important clues to understand the physical processes of AGNs. Below we explain these physical processes detailedly in each waveband:

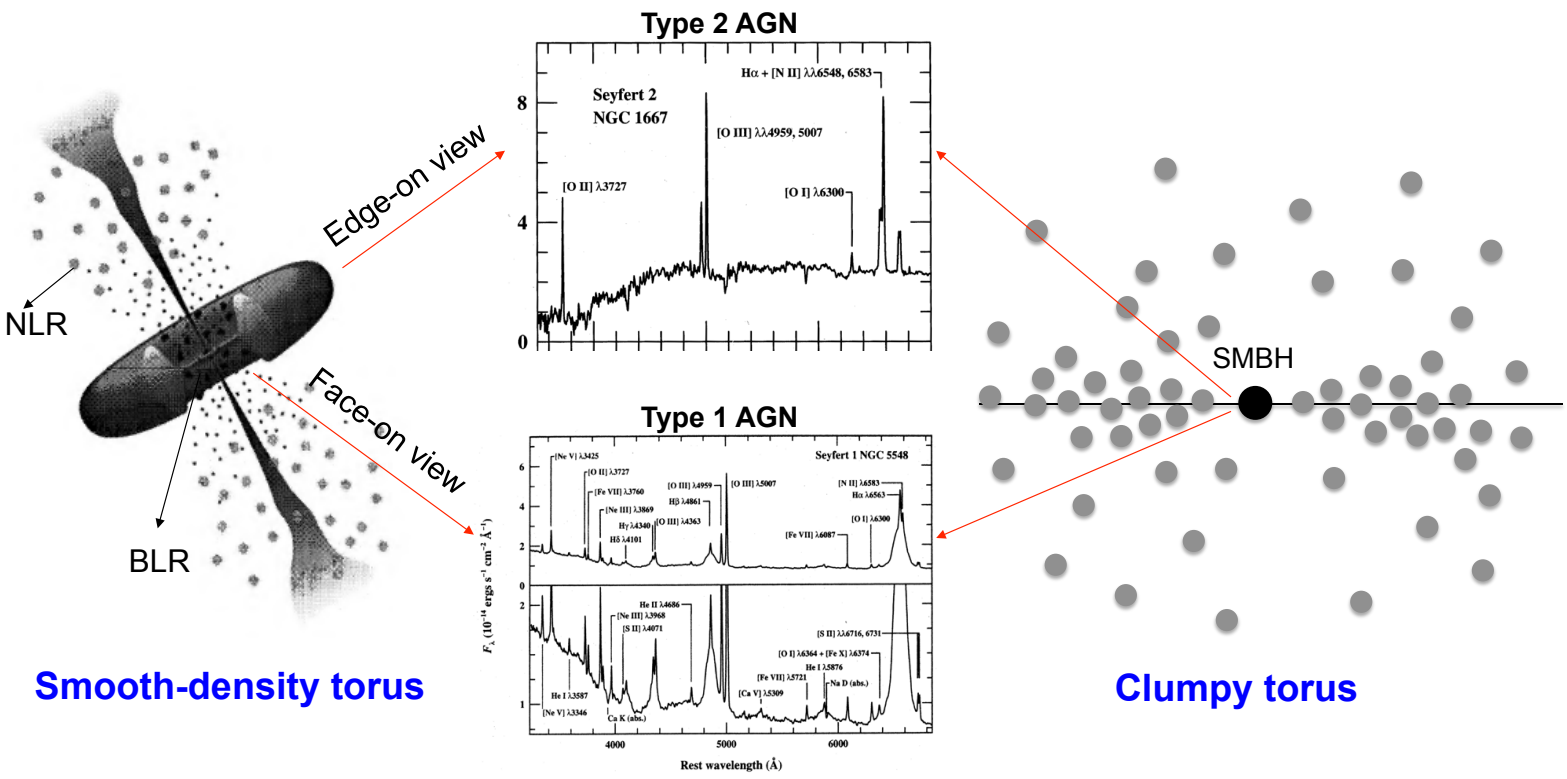


Figure 1.1: Illustration of the optical spectra of Type 1 and Type 2 AGNs according to the unified models with smooth-density torus (left) and clumpy torus (right). The left scheme of the AGN unified model is taken from Urry & Padovani (1995); in the middle are optical spectra of type 1 (bottom) and type 2 (top) AGNs taken from Peterson (1997). In the right is the clumpy torus model adapted from Nenkova et al. (2008b).

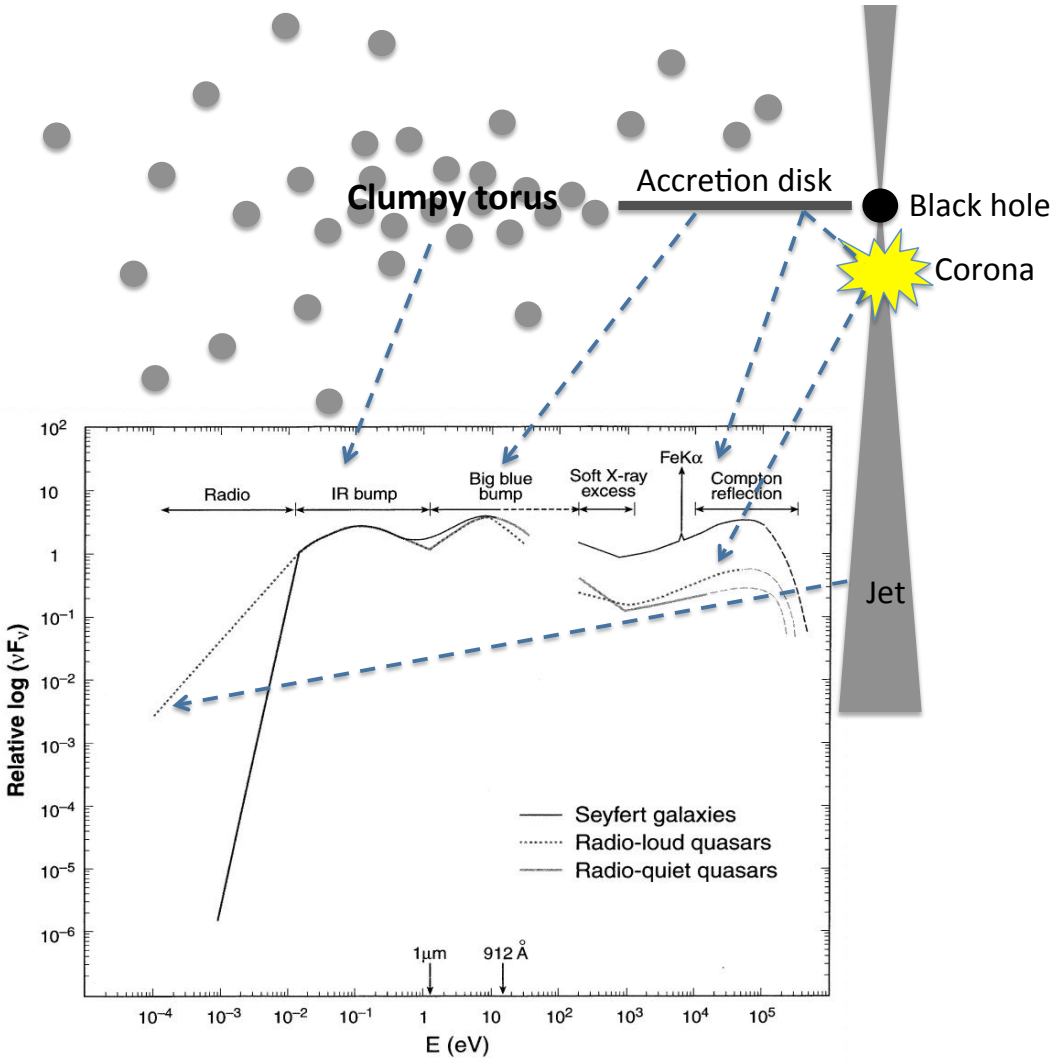


Figure 1.2: AGN spectral energy distribution from radio to X-ray. The subplot showing AGN continuum is adapted from Koratkar & Blaes (1999).

• **X-ray**

The X-ray emission includes two major components: (a) a primary power-law continuum; (b) and a reflection component. The intrinsic X-ray continuum is emitted from the hot corona or jet, showing a power-law SED extending from ~ 1 keV to over 100 keV. The power-law continuum can be fit by the form: $F_\nu \propto \mu^{-\Gamma+1} \propto \nu^{-\alpha}$, where Γ is called the "photon index", and α is referred to the "energy index". For radio-loud AGNs, the spectra are generally flatter than radio-quiet AGNs because of the additional radiation emitted by inverse-Compton scattering of the electrons in the jet. For the "soft excess" between 0.2 keV and 1 keV, the physical origin is not yet clear. It could be an extension of the big blue bump or produced from warm gas located in the accretion disk or BLR.

The reflection component is attributed to the reflection of the primary continuum by

Compton scattering in the accretion disk or the cold clumpy gas around (e.g., clouds in BLR or torus). This Compton reflection contains a continuum with a peak at ~ 30 keV, which is called "Compton hump" in the X-ray spectrum. In addition to the continuum, the Compton reflection also produces fluorescent emission lines (Reynolds, 1999). The iron Kα (Fe Kα) line at 6.4 keV is the most dominant feature in the X-ray spectrum that presents not only a narrow line from farther regions (e.g. torus or NLR), but also a broad line component. Many works have found that the Fe Kα line shows a very broad and asymmetric profile, which is supposed to be the consequence of relativistic Doppler shift and gravitational redshift (Tanaka et al., 1995; Nandra et al., 1997; Reynolds, 1997). This broad Fe Kα line profile is believed to be reflected from the innermost part of the accretion disk, and provides a powerful probe for understanding the conditions near SMBHs.

• Ultraviolet/optical

The UV and optical continuum of a type 1 AGN is attributed to the thermal emission from the accretion disk, and the spectrum can be approximately fit with a power law. The temperature of the accretion disk is in the range of $10^5 - 10^6$ K with a maximum energy flux at wavelength of ~ 100Å obtained by the Wien's displacement law (eq. 1.1):

$$\nu_{max} = \frac{2.8kT}{h} \approx 3.6 \times 10^{16}\mathrm{Hz} \tag{1.1}$$

This leads as a consequence to the "big blue bump" in the UV/optical continuum. The "soft X-ray excess" might be the high-energy end of this feature.

In addition to the "big blue bump", broad emission lines are also one of the prominent characteristics in the UV/optical spectra of type 1 AGNs (see the spectrum in the bottom right of Figure 1.1). These lines are produced by photoionization in the BLRs of high-density clouds with gas density $\geq 10^9\,\mathrm{cm}^{-3}$. The width of the broad lines is generally in the range between$\sim 500\,\mathrm{km\,s}^{-1}$ and $\sim 10000\,\mathrm{km\,s}^{-1}$, with a typical value of $5000\,\mathrm{km\,s}^{-1}$ (Peterson, 1997). This broadened line profile is due to the Doppler shift caused by the bulk orbital motion in the BLR, and can be used as a probe for understanding the properties of the black holes (e.g., black hole mass; Krolik, 2001; Peterson et al., 2004).

• Infrared

The main characteristics of the IR continuum are : (a) a minimum at $\sim 1 - 2\mu$m, and (b) an "IR bump" at $\sim 10 - 30\mu$m. The 1μm minimum corresponds to the sublimation temperature of dust grains (usually between 1000 to 2000 K). The IR bump is due to the thermal emission of dust with a temperature between 50 and 1000 K. The origin the IR emission is still controversial. The continuum could be emitted by hot dust (< 2000K) in the nuclear regions, or by the mechanism of reprocessing of the AGN primary emission, such as big blue bump.

• Submillimeter/radio

The main emission for the radio/sub-millimeter continuum is non-thermal synchrotron radiation emitted by relativistic electrons in the jet. The radio/sub-millimeter continua

have big differences between radio-loud and radio-quiet AGNs. For radio-loud objects, a continuum extends from the radio to the far-IR with a flatter spectrum shape. However, in a radio-quiet object, the spectrum decreases sharply from the far-IR, through the sub-millimeter to longer wavelengths ($F_\nu \propto \nu^\alpha$, $\alpha > 3$). This feature is so-called "the sub millimeter break".

1.1.3 Properties of obscured/type 2 AGNs spectra

In practice, a large fraction of AGNs are obscured in the Universe and appear faint in optical. For these optically low-luminosity AGNs, the utilization of X-ray, IR, and radio wavebands provide alternative ways to select AGNs. Martínez-Sansigre et al. (2006) used IR and radio criteria to select high-redshift type 2 quasars at $z \geq 2$. They claimed that the type 2 to type 1 AGN ratio is $\sim 3:1$. Fiore et al. (2009) combined deep X-ray data from *Chandra* with IR data from *Spitzer*/MIPS to select heavily obscured Compton-thick quasars. They found that the fraction of all types of AGNs to the total MIPS source population is $49\% \pm 10\%$. Buchner et al. (2015) also found that obscured AGNs with $N_H > 10^{22} cm^{-2}$ account for 77% of the number density of the accreting SMBH population.

The AGN emission could be absorbed by dust or gas from the circumnuclear region or the host galaxy. Figure 1.3 gives a simple example scheme for the absorption from the torus and host. The level of absorption can be varied by the number, composition, and geometry of the absorbers along the line of sight(e.g., Nenkova et al., 2008a,b; Elitzur, 2008; Buchner et al., 2015). Here we briefly describe the main features of the obscured/type 2 AGN spectrum.

- **X-ray**

For obscured AGNs, depending on the neutral hydrogen column density (N_H), the X-ray emission is absorbed via two mechanism: (1) photoelectric absorption and (2) Compton scattering. With $10^{22} < N_H < 1.5 \times 10^{24} cm^{-2}$ (Compton thin), the obscuration is caused by photoelectric absorption which dominates below 3 keV. With $N_H > 1.5 \times 10^{24} cm^{-2}$(Compton thick), the absorption is dominant from 7 keV to 30 keV and is caused by Compton scattering. Figure 1.4 shows how the X-ray emission decreases with increasing values of N_H, which represents the amount of gas along line of sight.

- **UV/optical/IR**

In type 2 AGN, the optical to UV continuum is absorbed by dust and the radiation is re-emitted at longer wavelengths. Therefore the spectrum is dominated by the stellar radiation from the host galaxy, and the big blue bump feature observable in type 1 AGN is not present in the type 2 AGN continuum. In addition, the optical spectrum only shows the narrow emission lines from the NLR, whereas the broad emission lines are absent because of the torus orientation effect that hides the BLR from the line of sight (see the spectrum in the upper right of Figure 1.1). The NLR has a relatively low density (i.e., $\sim 10^3\,cm^{-3}$), producing forbidden lines such as [OIII]λ5007 or [NII]λ6583. The width of narrow lines

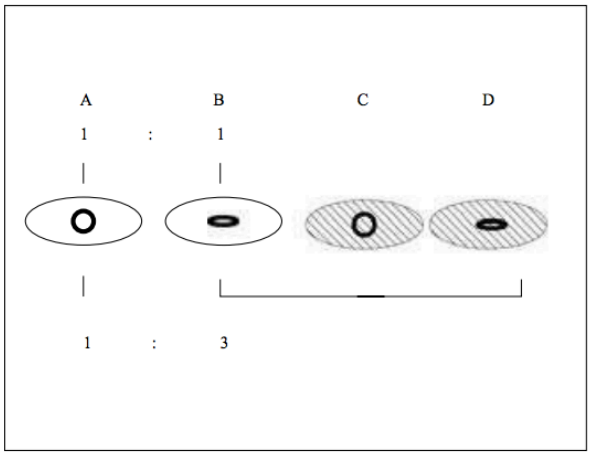


Figure 1.3: Schematic representation of the unobscured (type A) and obscured (type B, C, D) quasars. The inner black rings represent the torus surrounding the SMBH with a face-on (type A and C) or edge-on (type B and D) view. The shaded ellipses represent the dust in the host galaxy. The figure is from Martínez-Sansigre et al. (2006) who claimed that the type 2 to type 1 AGN ratio is $\sim 3:1$.

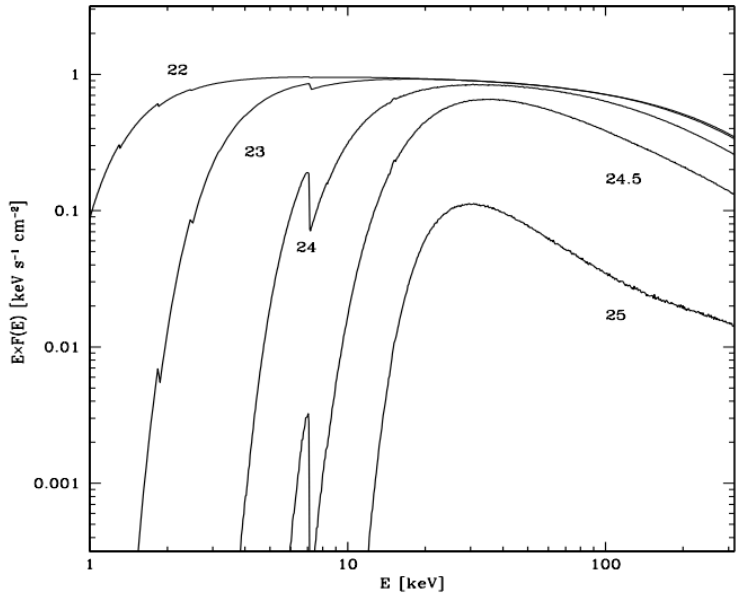


Figure 1.4: Typical obscured AGN X-ray spectrum showing the effects of photoelectric absorption and Compton down-scattering (a power law with photon index $\Gamma = 2$). Labels correspond to the logarithm of N_H (figure from Comastri, 2004).

usually less than $\sim 500\,\mathrm{km\,s^{-1}}$, with typical values of $300 - 400\,\mathrm{km\,s^{-1}}$ (Peterson, 1997).

Since star-forming galaxies also emit narrow lines, we can use their line ratios (e.g., [OIII]λ5007/Hβ) to distinguish type 2 AGNs from star-forming galaxies. However for sources without spectroscopic data, it is difficult to tell the difference between type 2 AGNs and star-forming galaxies only by the UV/optical continua.

For the IR continuum, in addition to the dust radiation from the torus and the host, the spectrum also contains a large contribution from the reprocessing of optical to UV continuum. Therefore, in comparison with type 1 AGN continua, type 2 AGN continua show more flux in the IR waveband (Haas et al., 2003).

• Sub-mm/Radio

Because Sub-mm/Radio emission is not absorbed by dust or gas, the spectrum of type 1 and type 2 AGNs are similar in sub-mm/radio. Thanks to this characteristic, we can search for the heavily obscured AGNs that are missing in optical and soft X-ray wavebands using sub-mm/radio observations, as Martínez-Sansigre et al. (2006) have demonstrated.

1.2 AGN evolution

The growth history of SMBHs can be reconstructed from the observations at low and high redshift. In this section, we first present the works from the multi-wavebands observations at lower redshifts for constraining the SMBH growth in the local universe. Then we present the recent results at higher redshifts that have implications for the AGN and galaxy (co)evolution in the early Universe.

1.2.1 BH growth in the local Universe

In the local Universe, it has been well established that most galaxies harbor a SMBH in the center (Kormendy & Richstone, 1995; Magorrian et al., 1998). Since SMBHs mainly grow through matter accretion during the AGN phase (e.g., Soltan, 1982; Rees, 1984; Marconi et al., 2004), the fraction of galaxies with a SMBH in an active phase provides a measurement of the time scale of BH growth (i.e., AGN duty cycle). For example, Goulding et al. (2010) measured the AGN duty cycle for the IR-bright galaxies ($L_{\mathrm{IR}} > 3 \times 10^9 \mathrm{L}_\odot$) with low X-ray luminosities (average $L_{\mathrm{X}} < 10^{43} \mathrm{L}_\odot$). They found that the fraction of these galaxies is $\sim 10 - 30\%$ over the M_{BH} range $\sim 10^6 - 10^8\,\mathrm{M}_\odot$. In Figure 1.5, Alexander & Hickox (2012) compared the result of Goulding et al. (2010) with Best et al. (2005) who showed a lower faction ($\sim 5\%$ over the $M_{\mathrm{BH}} \sim 10^{6.5} - 10^8\,\mathrm{M}_\odot$) of galaxies hosting optically luminous AGNs (defined as $L_{[\mathrm{OIII}]} > 3 \times 10^7\,\mathrm{ergs^{-1}}$). In contrast to the approximately constant fraction of IR-bright and optical-bright sources (crosses and filled squares in Figure 1.5, respectively), the fraction of galaxies with luminous radio AGN ($L_{1.4\,\mathrm{GHz}} > 10^{24}\mathrm{W\,Hz^{-1}}$) shows a strong dependence on the black hole mass, increasing from $\sim 0.01\%$ to 10% with M_{BH} from $10^{7.5}\,\mathrm{M}_\odot$ to $10^9\,\mathrm{M}_\odot$).

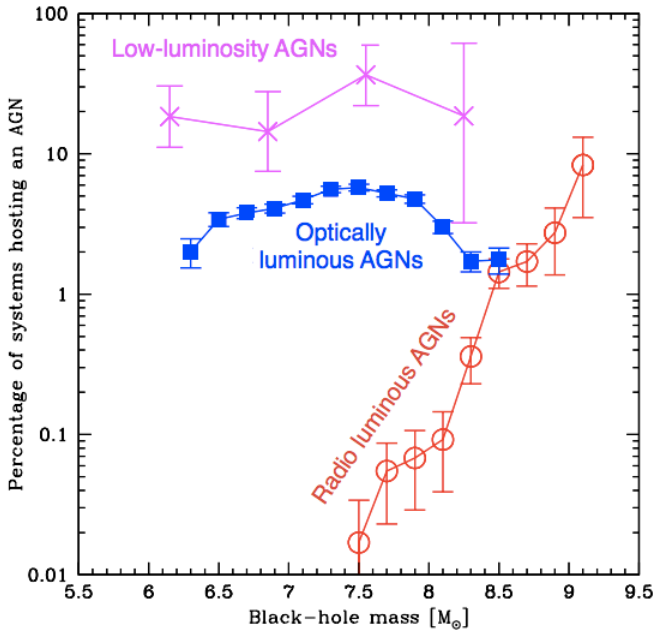


Figure 1.5: AGN fraction as a function of BH mass for different AGN selections at $z < 0.3$. Crosses: IR-bright sample from Goulding et al. (2010); filled squares: optical-bright sample from Best et al. (2005); open circles: radio luminous AGNs from Best et al. (2005). Figure taken from Alexander & Hickox (2012).

The AGN fraction is not only related to the BH mass, but also to the host stellar mass (M_*). For example, Xue et al. (2010) showed that the fraction of galaxies hosting luminous AGN ($L_{\rm X} > 10^{43} L_\odot$) at $z \sim 1$ increases from $\sim 0.3\%$ to 3% with M_* increasing from $\sim 10^{10}\,M_\odot$ to $\sim 10^{11}\,M_\odot$ (Figure 1.6).

However, the rate of BH growth cannot just be measured using the AGN fraction. The mass accretion rate ($\dot{M}_{\rm BH} \propto \epsilon L_{\rm Edd}$, where ϵ is the radiative efficiency) and Eddington ratio (defined as $\lambda_{\rm Edd} = L_{\rm bol}/L_{\rm Edd}$) are required to measure the growth rate of SMBHs. For a black hole growing for a given ϵ and an Eddington ratio $\lambda_{\rm Edd}$, the growth time from an initial mass M_{in} to a final mass M_{fin} is derived by (Volonteri, 2010):

$$t_{\rm growth} = 0.45\,{\rm Gyr}\frac{\epsilon}{1-\epsilon}\lambda_{\rm Edd}\ln\left(\frac{\rm M_{fin}}{\rm M_{in}}\right) \tag{1.2}$$

In the local Universe, the typical accretion time scales of the AGN phase is $\sim 10^7 - 10^9$ yr (e.g., Marconi et al., 2004), and the majority of AGNs have Eddington ratio of $\lambda_{\rm Edd} \sim 10^{-6} - 10^{-3}$ (e.g., Goulding et al., 2010). With an Eddington ratio of $\lambda_{\rm Edd} \sim 10^{-3}$, the growth time of a BH is comparable to the age of the Universe, which is not possible according to the studies of AGN duty cycle. This suggests that the SMBHs must have rapid growth in the past (Goulding et al., 2010; Alexander & Hickox, 2012).

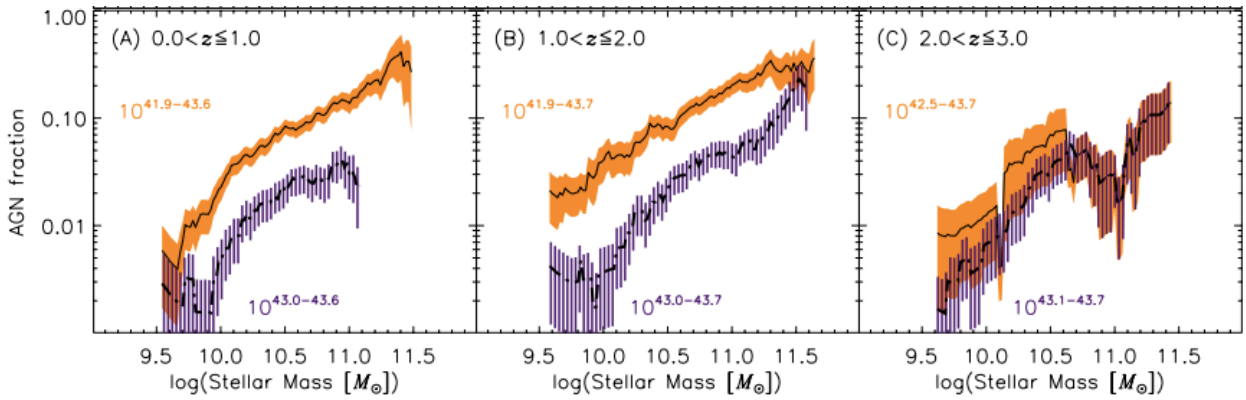


Figure 1.6: AGN fraction as a function of stellar mass in redshift bins. The orange shaded region indicates the 1 σ confidence range of the sources with $41.9 < \log L_{(0.5-8\,\mathrm{keV})} < 43.7\,\mathrm{erg\,s^{-1}}$, and purple points with uncertainties indicate the 1 σ confidence range of the sources with $43.0 < \log L_{(0.5-8\,\mathrm{keV})} < 43.7\,\mathrm{erg\,s^{-1}}$. Figure taken from Xue et al. (2010).

1.2.2 AGN evolution and BH growth in the distant Universe

The extensive data on SMBHs in nearby galaxies have provided a dramatic insight into the growth of BHs in the local Universe. However, to have an overall view of the history of BH accretion, we need to understand how BHs evolve with time, and how they grew in the early Universe. Studies of the luminosity function of optical quasars (bright AGNs with $L_{\mathrm{bol}} > 10^{45}\,\mathrm{erg\,s^{-1}}$) in 1960-80s first discovered the existence of AGN evolution (e.g., Schmidt, 1968; Schmidt & Green, 1983; Boyle et al., 1988). The density and luminosity evolution at $z > 2$ was not clear until the 1990s, when large spectroscopic surveys of high-z quasars were carried out. Surveys like the Two Degree Field (2dF) redshift survey and Sloan Digital Sky Survey (SDSS) have derived the optical luminosity function up to $z \sim 6$, although the sample size becomes small at higher redshift (Boyle et al., 2000; Fan et al., 2001, 2006; Richards et al., 2006). Figure 1.7 presents the overall evolution of luminous quasars using these two surveys. The space density of bright quasars peaks at $z \sim 2-3$ and exponentially declines toward lower and higher redshift. It is still uncertain whether this trend continues at $z > 6$. To achieve higher redshifts, deeper and higher-resolution data in NIR waveband, e.g., KMOS and MOSFIRE, are needed to constrain the results.

Nevertheless, optical AGN selection is only effective for bright and unobscured AGNs, and provides marginal constraints on obscured or low-luminosity AGNs. X-ray surveys are more efficient for selecting AGNs and provide more complete census of the AGN population, including both unobscured AGNs and obscured AGNs (note that some of the heavily obscured AGNs with N_{H} are still missing in X-ray observations). For instance, X-ray observations using *Chandra* and *XMM-Newton* have provided substantial number of X-ray surveys in both deep or wide fields, e.g., *Chandra* Deep Fields (CDF) (Alexander et al., 2003; Luo et al., 2008; Xue et al., 2010) and the Cosmological Evolution Survey (COSMOS;

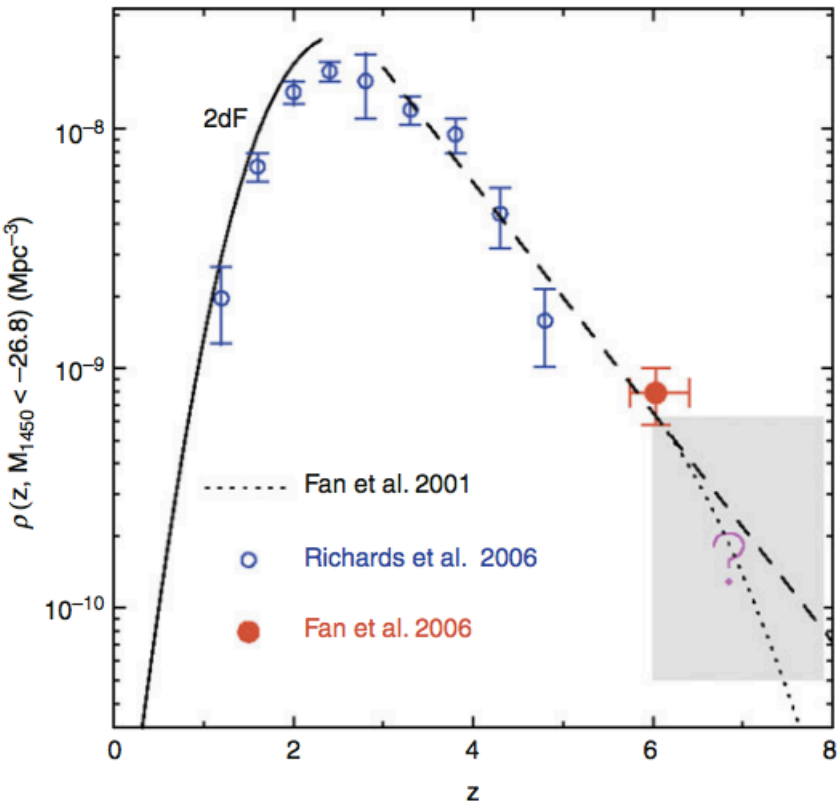


Figure 1.7: The space density of luminous quasars as a function of redshift. The solid line is the best-fit model from the 2dF at $z < 2.5$ (Boyle et al., 2000). The dashed line is the best-fitting model from the SDSS at $3.6 < z < 5$ (Fan et al., 2001). The blue open circles represent the result from Richards et al. (2006), and the red solid circles represent the result at $z \sim 6$ from Fan et al. (2006). Figure taken from D'Onofrio et al. (2012).

Cappelluti et al., 2009). These surveys provide more complete demographics of the AGN population.

In the X-ray luminosity function, lower X-ray-luminosity AGNs are found to peak at smaller redshift (as shown in Figure 1.8; Hasinger et al., 2005). This trend has been called "cosmic downsizing" which implies that more massive BHs underwent accretion processes in the distant Universe, whereas less massive BH did that in the nearby Universe (e.g., Cowie et al., 2003; Heckman et al., 2004). However the picture is still incomplete because we are missing a fraction of obscured AGNs. Recently many works have estimated such fraction, and found that it increases with redshift (e.g., Ueda et al., 2003; Hasinger, 2008; Aird et al., 2012; Buchner et al., 2015). This result implies that distant AGNs are surrounded by larger amounts of gas than AGNs in the local Universe. This raises questions: Where did the gas come from? Is there also at high redshift a connection between the amount of gas (and star formation rate) and BH accretion?

Many studies work on these topics and provide evidence for the connection between star

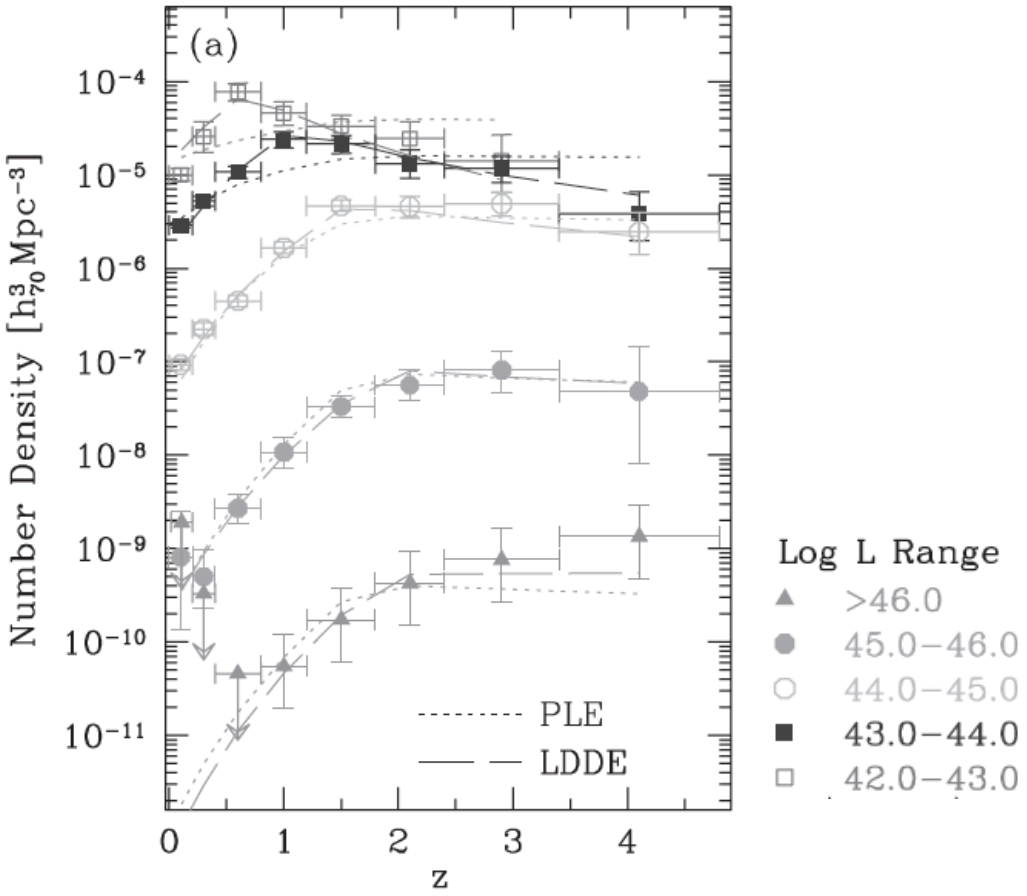


Figure 1.8: The number density of type 1 AGNs as a function of redshift in X-ray luminosity bins. Figure adapted from Hasinger et al. (2005).

formation and AGN activity at high redshift. Figure 1.9 shows the star formation history curve (Madau & Dickinson, 2014) which is similar to the evolution of luminous AGN space density presented in Figure 1.7, both with a peak at $z \sim 2$. This suggests that black hole growth and star formation are closely linked at all redshifts (Merloni et al., 2004; Silverman et al., 2008b; Madau & Dickinson, 2014). Silverman et al. (2008b) converted the X-ray luminosity to a mass accretion rate on to SMBH and compared with star formation history. They indicated that a global decline in star formation rate and black hole mass accretion rate over the last 8 Gyr may be an evidence of AGN feedback which has an impact on the host galaxy and quench the star formation activity. As shown in Figure 1.10, the SMBH mass growth rates derived from X-ray and Infrared data are compared with the cosmic star-formation rate density (Madau & Dickinson, 2014).

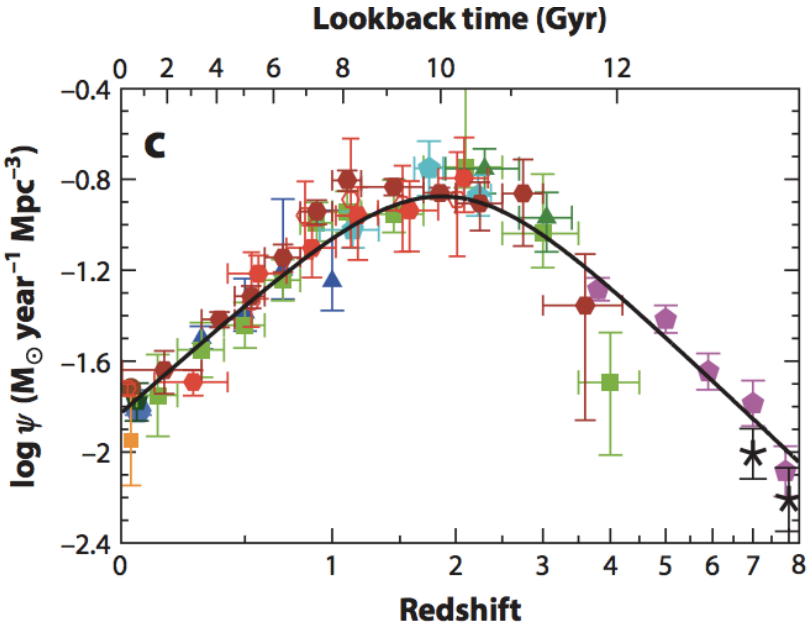


Figure 1.9: The cosmic star formation history. Figure taken from Madau & Dickinson (2014).

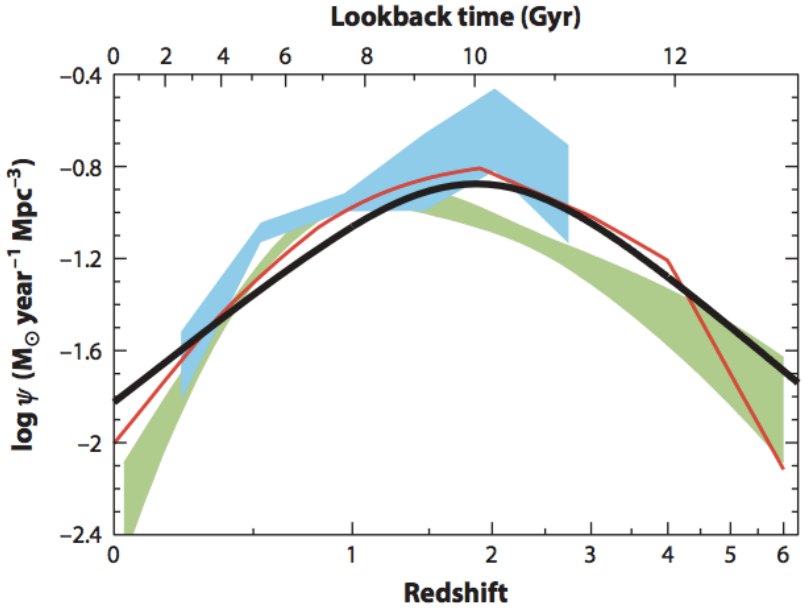


Figure 1.10: Comparison of the star formation history (black solid line) with the mass accretion rate on to SMBHs from X-ray (red curve and green shading) and Infrared (light blue shading) data scaled up by a factor of 3300. The shading indicates the $\pm 1\sigma$ uncertainty range. Figure taken from Madau & Dickinson (2014) .

1.3 Co-evolution of AGN and their hosts

The evolution of galaxy colors (and thus AGN host colors) can also be used for interpreting both galaxy evolution and how AGNs and galaxies influence each other. The studies are usually done by plotting the rest-frame color-magnitude diagram (CMD), originally using $U-V$ and M_V. In this plot, late-type galaxies with bluer colors due to dusty star-forming region and younger stellar population are well separated from early-type galaxies that have redder colors and older stellar population (e.g., Tully et al., 1998). Bell et al. (2004) first used the color-magnitude diagram showing this separation and introducing the concept of "blue cloud" and "red sequence" (Figure 1.11). They also used the diagram to study the galaxy evolution by investigating how the color distributions of early-type and late-type galaxies (i.e. passive or star forming galaxies) change with redshift. They revealed that the peak of the blue star-forming galaxies becomes redder toward the present day, and the number density of these blue galaxies drops sharply since $z \sim 1$. However, the interpretation of galaxy colors can be misled by dust reddening. Brammer et al. (2009) indicated that the galaxy colors in the green valley (i.e., intermediate region between blue cloud and red sequence) are largely affected by dust extinction. They have showed that correcting for dust reddening makes the separation between red and blue galaxies much more clear.

The CMD applied to AGN hosts can provide a direct comparison with normal galaxies distribution and evolution with the cosmic time. Several works have been done it at $z \sim 1$ in small survey areas, e.g., the EGS and CDFS, where bright AGNs are few and the majority of sources are moderate-luminosity AGNs. In these fields, the X-ray luminosity of the AGNs have a weak correlation with their optical emission, and thus have host-dominated SEDs(e.g., Nandra et al., 2007; Silverman et al., 2008a; Xue et al., 2010). These works, purely based on spectroscopic samples limited to specific redshift ranges, found no apparent bimodality that instead is clearly presented for inactive galaxies (Figure 1.12). In addition, these studies also found a large fraction of X-ray selected AGNs placed in the green valley, which is considered as an evolutionary transition from blue cloud (star-forming galaxies) to red sequence (passive galaxies). This result suggests that AGN feedback may contribute to AGN evolution during the phase of green valley where host star formation are suppressed by powerful AGN. However, Cardamone et al. (2010a) claimed that the majority of green-valley AGN hosts are intrinsically blue but reddened by dust. They found a clean separation between red and blue galaxies after applying dust de-reddening.

In fact, examining the relation between absolute magnitude and color is just a simply method of representing the more fundamental relation between stellar mass and star formation rate (Daddi et al., 2007; Rodighiero et al., 2011), as shown in Figure 1.13. The figure shows the main sequence of the star-forming galaxies, where the passive galaxy is located below it. Stellar massed and star-formation rates can be reliably derived for galaxies, as it is shown by use of simulation (see Mobasher et al., 2015; Santini et al., 2015), but it is not the case for galaxies hosting an AGN. Depending on the relative contribution and different type of AGNs and their hosts, the final SEDs of the source can vary dramatically. Strong star-forming objects will have strong emission lines and the same lines will be excited by

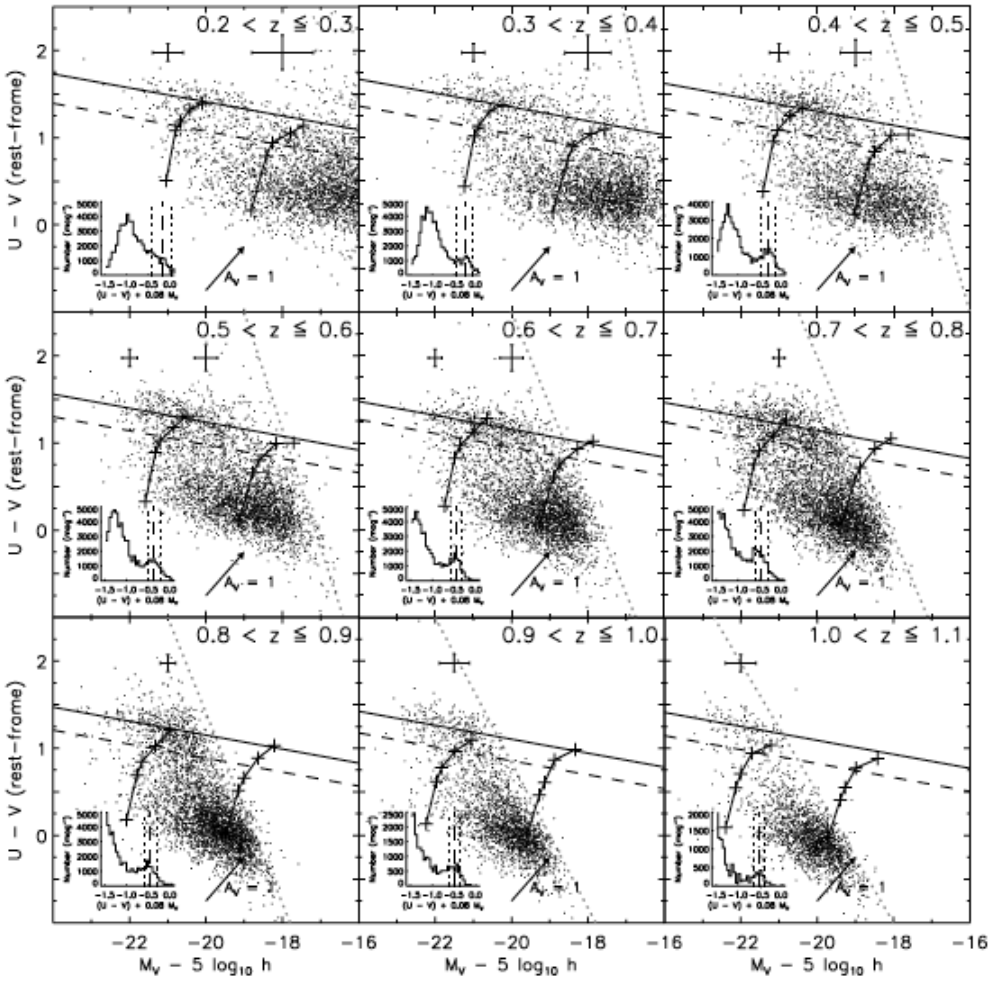


Figure 1.11: The rest-frame color-magnitude diagram (CMD) for the R-selected galaxies in the COMBO-17 at redshift range $0.2 < z < 1.1$. The solid line is the linear fit of the red sequence, and the dashed line is the cut between red and blue galaxies. The color distribution at each redshift range is inset In the left bottom of each plot. Figure taken from Bell et al. (2004).

an AGN. Both will contribute to the continuum, which is normally used for estimating stellar masses and star formation. Santini et al. (2012) and Bongiorno et al. (2012) tried to compute the stellar masses for AGNs using a different approach, while they approximately agree for the masses of low luminosity AGNs and both of them alert the reader to be suspicious of the masses computed for the bright AGNs. When comparing the values they obtained for the same bright sources, we find not only that the discrepancy is huge, but also the associated errors are large, making the use of the derived masses unreliable.

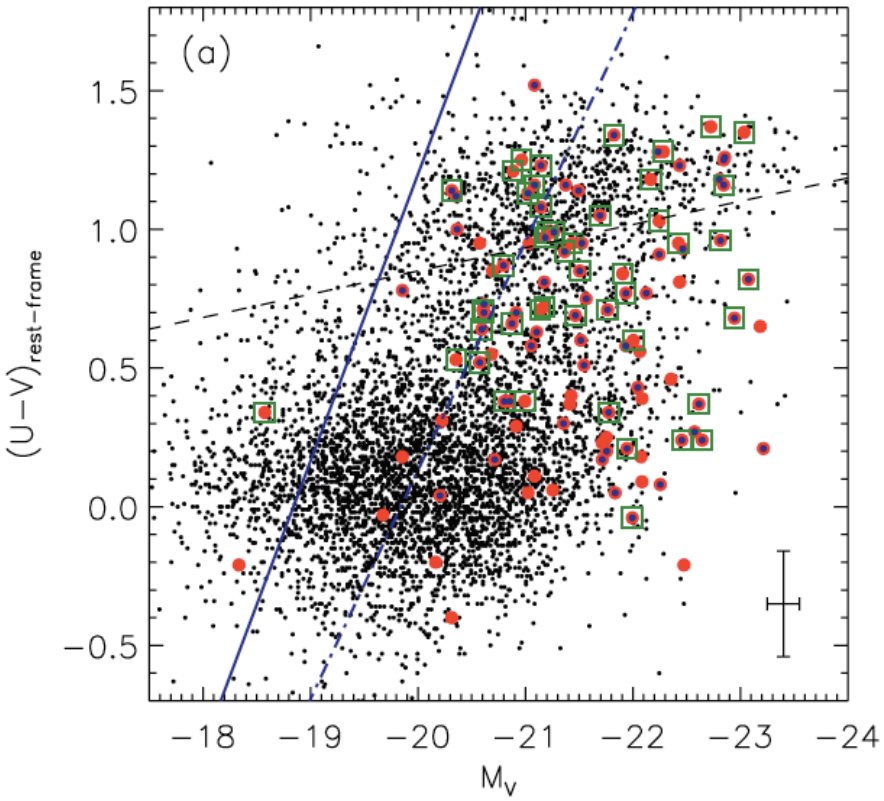


Figure 1.12: Rest-frame color-magnitude diagram of AGN host galaxy colors in the ECDFS within the redshift range $0.4 \leq z \leq 1.1$. Galaxies are shown in black dots while X-ray detected AGNs are displayed in red filled circles. Bulge-dominated AGN host galaxies are marked with blue filled circles. Open green squares indicate the AGNs with X-ray hardness ratios larger than 0.2. Figure taken from Silverman et al. (2008a).

1.4 Techniques for selecting high-z sources

To have a more complete view on the AGN and galaxy (co)evolution, it is necessary to extend the studies to higher redshifts. The investigation on the high-z sources can provide us hints on how black holes form and growth in the early Universe and how AGNs contribute to the structure formation to the Universe. For example, the study of the X-ray luminosity function for high-z AGNs (as we discussed in Section 1.2) can give us a direct perspective on the evolution of AGN activity with cosmic time. Due to the observational limits, the spectroscopically confirmed source population is getting incomplete at increasing redshift.

Historically a number of alternative methods have been developed for filling this lack of sources, like (a) NIR-selected distant red galaxies (e.g., DRG, Franx et al., 2003), (b) extremely red objects (e.g., ERO, Elston et al., 1988), (c) the BzK color-color selection (Daddi et al., 2004), and (d) the Lyman break galaxy (LBG) technique (Steidel et al., 1996,

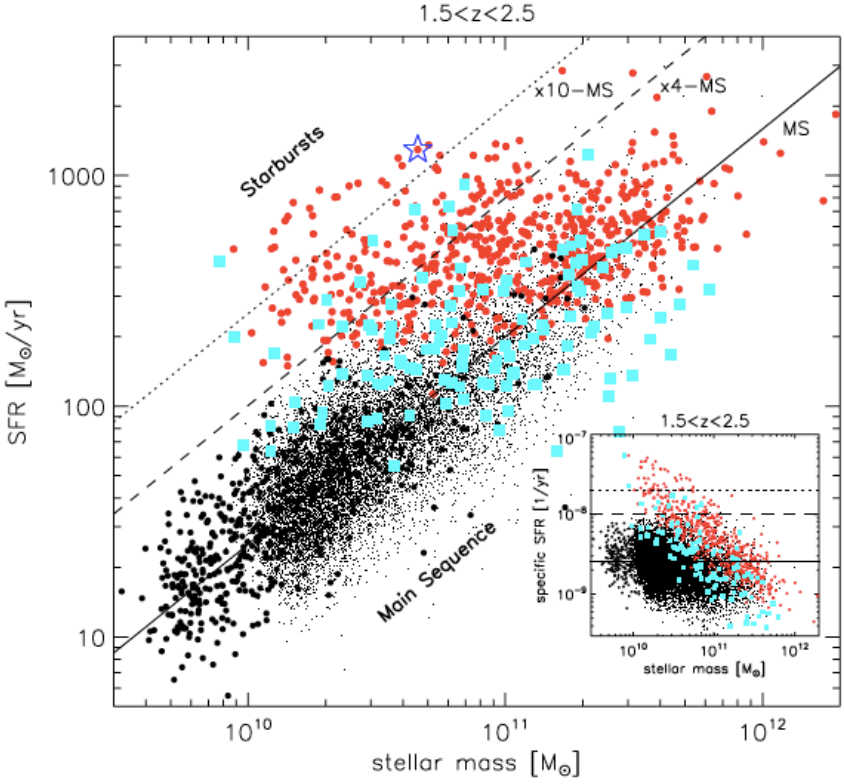


Figure 1.13: Stellar mass-star formation rate relation at $1.5 < z < 2.5$. Red filled circles and cyan squares represent the sources from the PACS-COSMOS and BzK-GOODS, respectively. The black dots indicate the BzK-COSMOS sources. The solid black line indicates the main sequence (MS) for star-forming galaxies at $z \sim 2$ defined by Daddi et al. (2007). Figure taken from Rodighiero et al. (2011).

1999, 2003). They all use the same principle, namely use specific bands to identify typical features in the SEDs (e.g. Balmer break, Lyman break) at the redshift of interest. For instance, EROs are selected at $z \geq 1$ by the NIR color $(R-K)_{vega} > 5$ or $(I-K)_{vega} > 4$. This extremely red color could come from old stellar populations in early-type and S0 galaxies, or from dusty starburst galaxies. Moriondo et al. (2000) studied the morphology of EROs and found that $\sim 50-80\%$ are indeed elliptical galaxies and $\sim 15\%$ are irregular and/or interacting galaxies. DRGs are chosen at $z \geq 2$ by very red NIR colors with $(J-K)_{\rm vega} > 2.3$. This method was first defined by Franx et al. (2003) using deep VLT/ISAAC data. The SEDs of these galaxies is characterized by the Balmer break which is shifted to the J-band at $z \sim 2$. As shown in Figure 1.14, the $(J-K)_{\rm vega}$ color changes sharply around $z \sim 2$, and can be an effective criterion to select sources at $z \geq 2$. The same Balmer break feature is used in BzK color-color technique for selecting star-forming and passive galaxies at $1.4 \leq z \leq 2.5$ using z and K bands. The color criteria have been

tested empirically using the spectroscopic data from the K20 survey ($K < 20$) and justified by the simulations of active and passive synthetic stellar population. For sources at $z \sim 3$, the LBG technique is traditionally adopted to identify the high-z sample by the Lyman break feature. This method uses U_nGR color criteria to characterizes the UV continuum emitted from the star-forming galaxies. We summarize these well-known color techniques in Table 1.1.

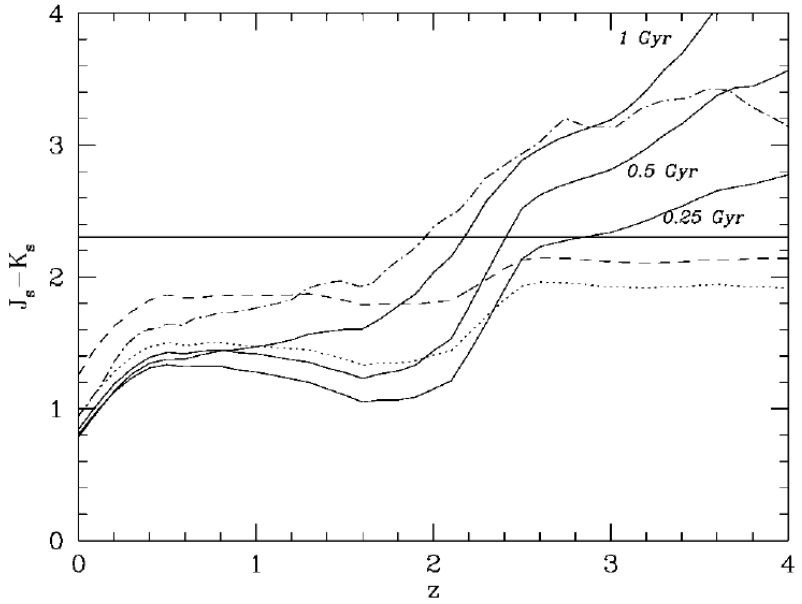


Figure 1.14: Color as a function of redshift for galaxies. The solid lines indicate single-age stellar population with age of 0.25, 0.5, and 1 Gyr. The dotted and dashed lines represent the stellar population with ages and reddening of 1 Gyr, $E(B-V) = 0.15$ and 100 Myr, $E(B-V) = 0.5$, respectively. Figure taken from Franx et al. (2003).

Table 1.1: Color selections of high-z sources

Method	Criteria	Redshift range	Nature	Reference
Extremely Red Objects (ERO)	$(R-K)_{\rm vega} \geq 5$	$z \geq 1$	Old elliptical galaxy or dusty star-forming galaxies	Elston et al. (1988)
Distant Red Galaxy (DRG)	$(J-K)_{\rm vega} > 2.3$	$z \geq 2$	Old elliptical galaxy or dusty star-forming galaxies	Franx et al. (2003)
BzK selection	$BzK = (z-K)_{\rm AB} - (B-z)_{\rm AB}$	$1.4 < z < 2.6$	Star-forming: $BzK > -0.2$ Passive: $BzK < -0.2$ and $(B-z)_{\rm AB} > 2.5$	Daddi et al. (2004)
Lyman Break Galaxy (LBG)	$(G-R)_{\rm AB} < 1.2$ $(U_n-G)_{\rm AB} > (G-R)_{\rm AB} + 1.0$ $19 < R_{\rm AB} < 25.5$	$z \sim 3$	Rest-frame UV is absorbed by IGM, or by OB star in star forming galaxies	Steidel et al. (1996) Steidel et al. (2003)

These techniques were all originally tuned for non-active galaxies, however some of them are also applied to AGNs either for selecting high-z sample or for understanding their physical properties. For example, to make AGN population more complete for analyzing the

X-ray luminosity function, LBG technique is used to select high-z AGNs in Nandra et al. (2005) and Aird et al. (2008, 2010). Moreover, Rangel et al. (2013) used BzK technique to investigate the properties of X-ray selected AGNs. They found that passively evolving BzK galaxies have higher fraction of AGNs than star-forming BzK galaxies, which contradict the idea of the coevolution between BH growth and star formation at $z \sim 2$. In Brusa et al. (2005), they used ERO selection to study the X-ray obscured quasars and found that majority of EROs are powered by the accretion activity of SMBH.

Due to large uncertainties caused by faint photometries, it is necessary to carefully select high-z sample. Aird et al. (2008) and Aird et al. (2010) pointed out the importance of using redshift probability distribution function, i.e., $P(z)$, to select high-z sample. They simulated the function of $P(z)$ for high-z sample selected by LBG method, applying an additional weighting factor, w. The measurement of w considers the probability that a source is located within a certain redshift range, and reduces the weight outside the target redshift range of LBG (see the detailed description in Section 5.3). This technique can provide effective number at high redshift and make the luminosity function more realistic.

In this thesis, we will use the $P(z)$ directly from our SED-fitting results of photo-z computation for galaxies and AGNs, and apply the value of w to discuss the efficiency of $P(z)$-technique and color methods. In the following section, we will demonstrate the challenges we are facing in the processes of running the SED-fitting.

1.5 The challenges that we are facing

For correctly modeling galaxy evolution, the availability of accurate redshifts for both normal galaxies and AGN is fundamental. Although redshifts measured via spectroscopic observations are very reliable, it is a problem to obtain large number of faint or high-redshift sources. Long exposure times are required for the faint sources typically found in deep field observations, and the relatively low sky density of AGN means that it is difficult to obtain large samples. Furthermore, for high-redshift sources, spectroscopic observations have observational limits such as: (1) the redshift range available to optical spectrographs and (2) the presence of the telluric OH lines that make it difficult to assign correct redshifts in near-infrared spectrographic observation from the ground. This restricts the availability of spectroscopic redshifts (spec-z), in particular for deep pencil-beam surveys. About 65% of sources in the Cosmic Assembly Near-IR Deep Legacy Survey (CANDELS; Grogin et al., 2011; Koekemoer et al., 2011) in the GOODS-S region are fainter than $H = 25$, beyond any reasonable spectroscopic limit. Similarly, only about 60% of the X-ray sources in the 4 Ms Chandra Deep Field-South (4Ms-CDFS) survey have reliable spec-z (Xue et al., 2011). Therefore, a large number of accurate photometric redshift (photo-z) are needed, particularly at the faint and high-redshift ends of the source distribution.

For normal galaxies, previous work has achieved photo-z accuracy (defined as $1.48 \times$ median($\frac{|\Delta z|}{1+z_s}$ of ~ 0.01) by using well-verified spectral energy distribution (SED) templates for galaxies in many fields (Ilbert et al., 2009; Cardamone et al., 2010b). Within the

Extended Chandra Deep Field-South (ECDFS), photo-z for many samples are available in the literature (e.g., Zheng et al. 2004; Grazian et al. 2006a; Wuyts et al. 2008; Cardamone et al. 2010b; Luo et al. 2010; Dahlen et al. 2013). Although the accuracy reported in each paper is similar, discrepancies emerge when comparing photo-z for objects without spectroscopic information, especially at high redshift and for faint sources. In the following sections we demonstrate the challenges we are facing in photo-z computation.

1.5.1 SED templates for AGNs

Photo-z accuracy depends on the number and resolution of wavelength bands available as already shown by Benítez et al. (2009). One of the fields with the greatest number of photometric bands is the Great Observatories Origins Deep Survey-South (GOODS-S; Giavalisco et al., 2004a), which has been observed recurrently as new facilities have become available. The GOODS-S region is at the moment in a unique niche as homogeneous and deep data (including the exquisite X-ray coverage with *Chandra*) are available. In addition to intermediate-band photometry from *Subaru* (Cardamone et al., 2010b) (Figure 2.3) and deep *Spitzer*/IRAC data (Damen et al., 2011; Ashby et al., 2013), *HST*/WFC3 NIR data from the CANDELS survey and J and K_S bands from the Taiwan ECDFS Near-Infrared Survey (TENIS; Hsieh et al., 2012) are now available. Deep NIR observations are necessary to obtain reliable redshifts at $z > 1.5$, where the prominent ~ 4000Å break shifts to NIR wavelengths. The availability of these new data will improve the accuracy of photo-z and push the good quality of photo-z to higher redshift.

The past decade has witnessed important developments in normal galaxy photo-z both by SED fitting and by machine learning techniques, and some of these improvements can be directly used for AGNs. For example, improvement of template-fitting photo-z by adding emission lines to the templates has been demonstrated by Gabasch et al. (2004, 2006) and Ilbert et al. (2006). Intermediate- and narrow-band (IB, NB) photometry is valuable to pinpoint emission lines in the SEDs (Ilbert et al., 2009; Salvato et al., 2009; Cardamone et al., 2010b; Matute et al., 2012), as simulations (e.g., Benítez et al., 2009) have predicted.

However, even with the best data, photometric redshifts for AGNs remain challenging (Salvato et al., 2009, 2011). Photo-z errors for AGNs can have a significant impact on galaxy/AGN coevolution studies. For example, Rosario et al. (2013) found that at high redshifts the AGNs tend to have bluer colors than inactive galaxies, implying younger stellar populations and higher specific star formation rates in the AGN hosts. This result, as Rosario et al. mentioned, may be biased by the spectroscopic selection effect and photo-z errors leading to a bluer host color. The photo-z failures are caused by observational limitations and improper templates used for photo-z computation. For example, Salvato et al. (2009, 2011) point out that due to the uncertainty of AGN/host galaxy contribution, we need to create proper AGN/galaxy hybrid templates to better model the AGN SEDs.

Unlike normal galaxies, SED of AGNs are more complex and need more tuning for photo-z computation. They could consist of mixed radiation emitted from the central AGN and the host, which can not be represented by pure galaxy or pure AGN templates. The variety of AGN SEDs is related to the AGN/host relative contributions, which can

be caused e.g. by observational limitations, orientation effect, and different types of AGNs in different evolutionary phases. In addition, templates for AGNs are often empirically obtained from nearby AGNs where the slit size allows sampling of only the nuclear region of galaxies. At high redshift, the size of entire galaxies are comparable to the size of slits, so that both AGN and host emissions are included in slits or fibers. Therefore, the empirical low-redshift templates do not represent the SEDs of high-redshift AGNs, and hybrids that include both AGN and host contributions are necessary.

Secondly, we assumed that the SEDs are vary according to the viewing angle along line of sight. For instance, Seyfert 1 AGNs show relatively higher emission in the bluer part of the continuum than Seyfert 2s which have more of a galaxy-like continuum (see Figure 1.1). Thirdly, the varying contribution from AGN and host could be the result of different physical properties in different evolutionary phases. Figure 1.15 (Haas et al., 2003) gives one idea of how the shapes of the AGN continuum (assumed face-on view) evolve with the scale of dust distribution. With larger-scale dust obscuration, the stellar light from host galaxies highly affects the spectral slopes from UV to optical, which appears red and galaxy-like. In the more "clear" quasar stage, the strong characteristic UV excess appears and the SED becomes AGN-dominated. The reasons above explain how complicated the AGN SED is and why we need to build representative AGN-Galaxy hybrids.

In addition, as shown in Figure 1.1, emission lines are prominent features in AGN spectra, arising either from gas ionized by the central nuclei and/or star-forming regions. As shown above, the available intermediate-band data help characterize line features, so including emission lines in our templates makes our hybrids even more powerful when computing photo-z.

1.5.2 Multi-wavelength associations

The situation for AGNs is further complicated by the need for an association with multi-wavelength data before a photo-z can be calculated. This makes the accuracy of positions for the X-ray sources and the method and data used for the associations of crucial importance. In this thesis the impacts of both issues are studied as they could result in different counterparts (and often multiple potential counterparts can not be excluded). This is demonstrated in Chapter 2, where we compare the X-ray detections from Xue et al. (2011) and Rangel et al. (2013) in CDFS, and Lehmer et al. (2005a) and Virani et al. (2006) in ECDFS. In both pair of works, the X-ray data were the same but the way in which the data were analyzed gave different detections. A challenge for this thesis was to assess when the sources identified from different methods were the same and when not.

The offset in X-ray coordinates, even marginal, can give an different association to the multi-wavelength database, especially in crowded areas. A simple match in coordinates between X-ray and other data is not feasible, particularly for X-ray surveys with a lower resolution than *Chandra* offered. But even at the resolution of *Chandra*, when NIR data are deep and resolved like the HST/WFC3 data in CANDELS, multiple associations are possible (see Section 2.2).

Several works have instead used the likelihood ratio method (e.g., Sutherland & Saun-

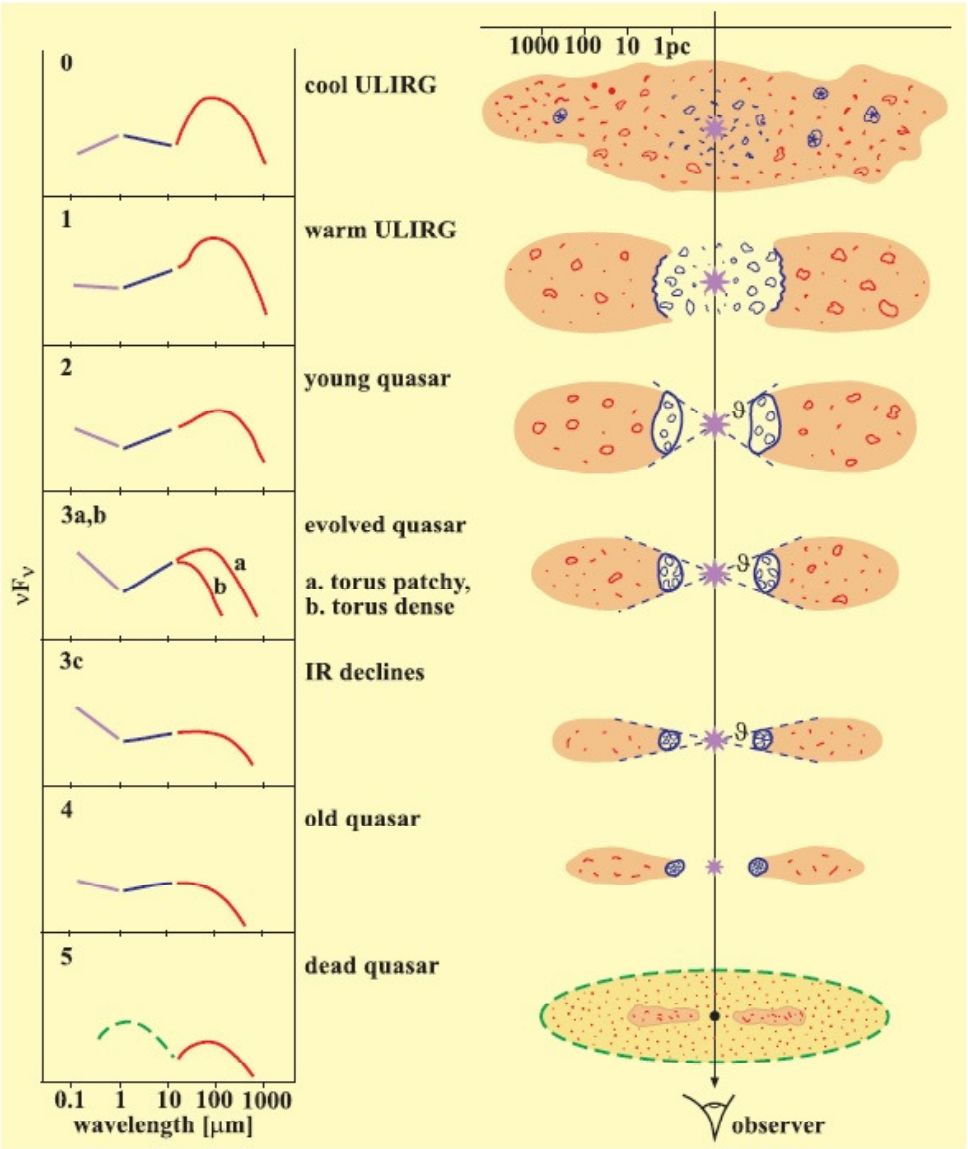


Figure 1.15: Scheme of quasar evolution for the dust distribution surrounding the AGN, with corresponding SEDs. Figure taken from Haas et al. (2003).

ders 1992; Brusa et al. 2007; Laird et al. 2009; Luo et al. 2010; Xue et al. 2011; Civano et al. 2012), which relies on homogeneous coverage in a given visible or infrared band. Most works repeat the association for several different reference bands and finally choose a counterpart by comparing the results. The likelihood approach is based on the distance of a source from the X-ray detection and the probability of correct association is purely based on the comparison between the magnitude distribution of the sources in the field and the sources closest to the X-ray detection. Even when the likelihood approach is used on several different reference bands, the association is then done independently, selecting as counterpart the most frequent. Especially for X-ray surveys with large positional uncertainties, this can affect the result. We adopted instead a new bayesian code (Salvato et al. 2015, in prep.) developed in the high energy group at MPE which by use of priors, trying to compute the correct association by working simultaneously on more than two catalogs.

1.6 The outline of this work

This thesis is organized as follows:

- Chapter 1 reviews the basic properties and mechanisms defining an AGN at various wavelengths. It also reviews our knowledge of AGN/galaxy (co)evolution in the local Universe and hints at the problems still open at high redshift. Finally, we introduces the difficulties that we will try to solve in this thesis.

- Chapter 2 introduces the fields and the multi-wavelength datasets we used in this work. We demonstrate how we combined various photometries from different surveys and how we defined the most likely optical and near-infrared counterparts to the X-ray sources.

- In Chapter 3, we present the photo-z results for normal galaxies and for AGNs. In particular we described the method of building a proper AGN-galaxy hybrid templates for SED-fitting in detail. Moreover we demonstrate how the photo-z accuracy improves dramatically when considering templates with emission lines and how the presence of Intermediate bands improves on the overall results.

- In Chapter 4, we investigate the rest-frame color magnitude diagram as a function of redshift for both galaxies and AGN hosts. We made correction for dust extinction and AGN contribution, discussing which factor affects the host color distribution more, and compare the results with previous works. Unlike those which where limited to $z \sim 1$, thanks to the accuracy of our photo-z and to the care we took in the AGN/galaxy decomposition, we can extend the analysis up to $z \sim 3$.

- In Chapter 5, we firstly study the traditional color methods like the Lyman break galaxy and BzK color-color selection. Then we test the robustness of our $P(z)$ technique (based on the SED-fitting result) by the measurements of the completeness and purity, comparing with the color methods. Finally we provide a high-z source list

at $z > 3$ with $P(z)$ information, comparing with high-z AGN sample available from Fiore et al. (2012) who defined high-z sources in the same area but with a different technique.

- In Chapter 6 we summarize the main results of the thesis, presenting an outlook of the next steps.

Throughout this work, we adopt the AB magnitude system and assume a cosmology with $H_0 = 70$ km s^{-1} Mpc^{-1}, $\Omega_\Lambda = 0.7$, and $\Omega_M = 0.3$ (Spergel et al., 2003).

Chapter 2

The fields: CANDELS/GOODS-S, CDFS, and ECDFS

The Chandra Deep Field South (CDFS) is an area of the sky with very low extinction by Galactic gas and dust. There are no bright stars ($M_{\rm V} \leq 14$) within 30′and the neutral hydrogen column density is very low ($N_{\rm H} \sim 8{\times}10^{19}{\rm cm}_{-2}$; Giacconi et al., 2002). This allows us to maximize the depth of X-ray observation, providing an opportunity to investigate the distribution and properties of supermassive black holes over cosmic time.

It is widely accepted that all AGNs in a field can not be identified using a single photometric band, which yields incomplete and partially overlapping subsample of the overall AGN population. A complete census of AGNs can only be achieved by combining the results from multi-wavelength selections. Nevertheless, a selection using X-ray band appears to be very efficient and rather complete for most classes of AGNs (Brandt & Hasinger, 2005). Therefore, in this work we identified AGNs using the X-ray detected sources from the *Chandra* space telescope.

Taking advantage of this field for extragalactic observations, the CDFS area has been observed repeatedly with a large variety of facilities and instruments. As a result, numerous datasets with different bands and depths are available depending on the exact location. In this work, we collected the multi-wavelength data from the X-ray to Infrared in the GOODS-S, CDFS and Extended-CDFS (ECDFS) fields. Reliable X-ray-to-optical associations and photometric redshifts can be obtained only when the data are homogeneous, and for this reason we split the area into subregions where the data are uniform. Three main regions share the same sets of data: Area 1 ($\sim$176 arcmin^2) is the region covered by CANDELS and GOODS-S, Area 2 ($\sim$ 290 arcmin^2) is the outer CDFS region surrounding CANDELS/GOODS-S, and Area 3 ($\sim$ 435 arcmin^2) is the ECDFS region outside the CDFS. Figure 2.1 shows the three regions.

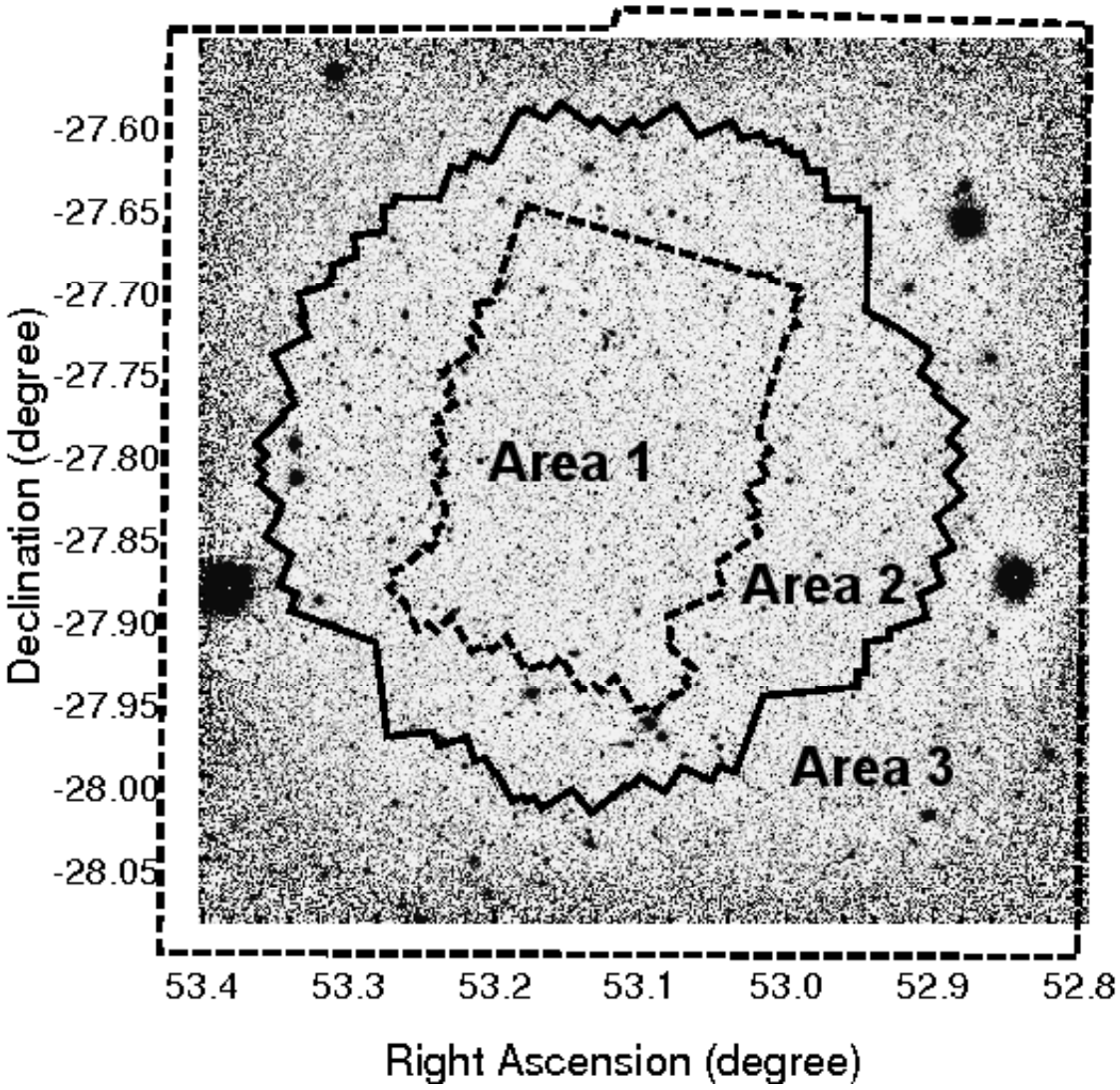


Figure 2.1: Major areas defined in ECDFS. Background is the negative $J + K_s$ image from TENIS. The inner dashed line encloses the CANDELS/GOODS-S area ("Area 1"), the solid line encloses the deep X-ray coverage (CDFS, "Area 2"), and the outer dashed line (ECDFS) shows the MUSYC (Cardamone et al., 2010b) coverage ("Area 3") that defines the full area used in this work.

2.1 Data set

2.1.1 X-ray data

The X-ray catalogs to cross-match were obtained from the *Chandra* survey of 4Ms-CDFS observations covering Area 1+2 and from the 250ks-ECDFS observations covering Area 3. Two independent groups (Xue et al., 2011; Rangel et al., 2013) have provided source catalogs for 4Ms-CDFS using different methods for data reduction and source detection. Similarly for Area 3, both Lehmer et al. (2005a) and Virani et al. (2006) have released X-ray source catalogs for the 250ks-ECDFS survey. We have cross-matched X-ray sources from both catalogs in each area.

For Areas 1+2 we used:

a. The 4Ms-CDFS source catalog of Xue et al. (2011) (hereafter X11) with 740 point-like X-ray sources. The sensitivity limits of the X-ray data are 3.2×10^{-17}, 9.1×10^{-18}, and 5.5×10^{-17} erg cm^{-2} s^{-1} for the full (0.5–8 keV), soft (0.5–2 keV), and hard (2–8 keV) bands, respectively.

b. The 4Ms-CDFS source catalog of Rangel et al. (2013) (hereafter R13)[1] produced using the analysis methodology of Laird et al. (2009). The catalog contains 569 point-like X-ray sources and has sensitivity limits 4.2×10^{-17}, 1.2×10^{-17}, and 8.8×10^{-17} erg cm^{-2} s^{-1} in the full, soft, and hard bands, respectively.

For Area 3 we used:

c. The 250ks-ECDFS X-ray catalog from Lehmer et al. (2005a) (hereafter L05) with 762 sources in the entire ECDFS of which 457 are in Area 3 (i.e., outside the 4Ms-CDFS area). Catalog sensitivity limits are 1.1×10^{-16} erg cm^{-2} s^{-1} in the soft (0.5–2 keV) band and 6.7×10^{-16} in the hard (2–8 keV) band.

d. The 250ks-ECDFS X-ray catalog from Virani et al. (2006) (hereafter V06) with 651 sources in the entire ECDFS of which 404 are in Area 3. Sensitivity limits are 1.7×10^{-16} and 3.9×10^{-16} erg cm^{-2} s^{-1} in the soft and hard bands, respectively.

2.1.2 Photometric data from UV to MIR

Altogether the CDFS region has been covered in up to 50 bands from ultraviolet (UV) to mid-infrared (MIR) as listed in Table 2.1 and Table 2.2. Table 2.3 summarizes the catalogs used in each area.

1. Area 1: In this region, we primarily used the CANDELS-TFIT multi-wavelength catalog of Guo et al. (2013, hereafter G13), which covers the CANDELS GOODS-S area with 18 broad-band filters (Figure 2.2) mostly from space observatories. The photometry was based on template-fitting (TFIT; Laidler et al., 2007), using the high resolution WFC3/H-band image to detect sources and define apertures, which were then used for

[1] The 4Ms-CDFS X-ray catalog of R13 is available under [Surveys] > [CDFS] through the portal `http://www.mpe.mpg.de/XraySurveys`.

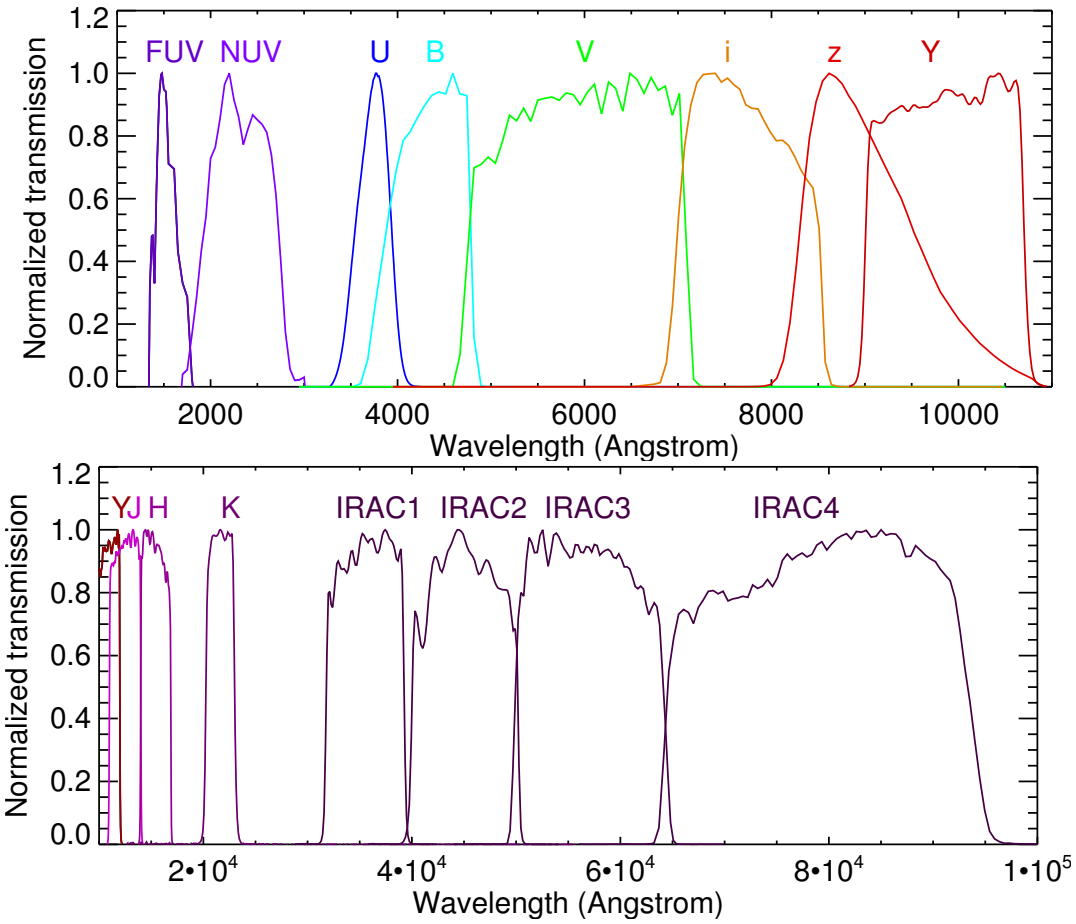


Figure 2.2: Optical (top) and NIR/MIR (bottom) broad-band filter transmission curves.

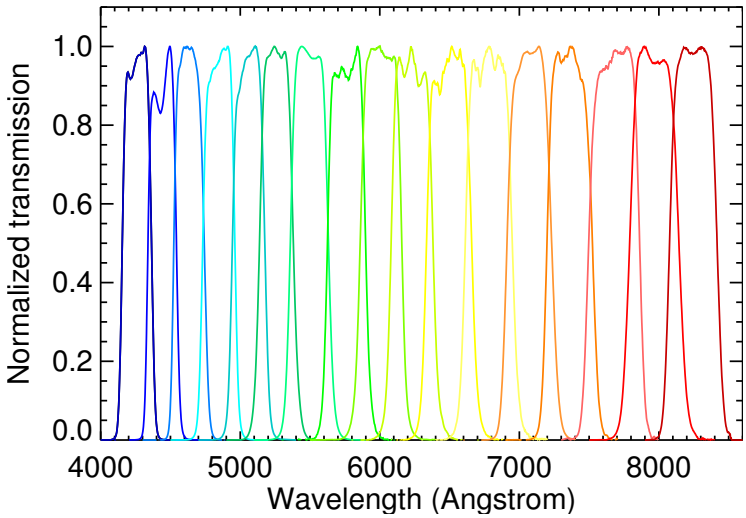


Figure 2.3: Subaru medium-band filter transmission curves.

Table 2.1: Photometric Data in $\text{TFIT}_{\text{CANDELS+IB}}$ catalog

Filter	λ_{eff} (Å)	FWHM (Å)	5σ Limiting Depth (AB mag)	Instrument Telescope	Area
FUV[e]	1543	228	25.69	GALEX	1
NUV[e]	2278	796	25.99	GALEX	1
U-CTIO[a]	3734	387	26.63	Blanco/Mosaic-II	1
U-VIMOS[a]	3722	297	27.97	VLT/VIMOS	1
F435W[a]	4317	920	28.95/30.55[f]	HST/ACS	1
F606W[a]	5918	2324	29.35/31.05[f]	HST/ACS	1
F775W[a]	7693	1511	28.55/30.85[f]	HST/ACS	1
F814W[a]	8047	1826	28.84	HST/ACS	1
F850LP[a]	9055	1236	28.55/30.25[f]	HST/ACS	1
F098M[a]	9851	1696	28.77	HST/WFC3	1
F105W[a]	10550	2916	27.45/28.45/29.45[g]	HST/WFC3	1
F125W[a]	12486	3005	27.66/28.34/29.78[g]	HST/WFC3	1
F140W[a]	13635	3947	26.89/29.84[h]	HST/WFC3	1
F160W[a]	15370	2874	27.36/28.16/29.74[g]	HST/WFC3	1
Ks-ISAAC[a]	21605	2746	25.09	VLT/ISAAC	1
Ks-HAWKI[a]	21463	3250	26.45	VLT/HAWK-I	1
3.6μm-SEDS[a]	35508	7432	26.52	Spitzer/IRAC	1
4.5μm-SEDS[a]	44960	10097	26.25	Spitzer/IRAC	1
5.8μm-GOODS[a]	57245	13912	23.75	Spitzer/IRAC	1
8.0μm-GOODS[a]	78840	28312	23.72	Spitzer/IRAC	1
IA427[d]	4253	210	25.01	Subaru	1
IA445[d]	4445	204	25.18	Subaru	1
IA464[d]	4631	216	24.38	Subaru	1
IA484[d]	4843	230	26.22	Subaru	1
IA505[d]	5059	234	25.29	Subaru	1
IA527[d]	5256	243	26.18	Subaru	1
IA550[d]	5492	276	25.45	Subaru	1
IA574[d]	5760	276	25.16	Subaru	1
IA598[d]	6003	297	26.05	Subaru	1
IA624[d]	6227	300	25.91	Subaru	1
IA651[d]	6491	324	26.14	Subaru	1
IA679[d]	6778	339	26.02	Subaru	1
IA709[d]	7070	321	24.52	Subaru	1
IA738[d]	7356	324	25.93	Subaru	1
IA768[d]	7676	366	24.92	Subaru	1
IA797[d]	7962	354	24.69	Subaru	1
IA827[d]	8243	339	23.60	Subaru	1
IA856[d]	8562	324	24.41	Subaru	1

[a] CANDELS-TFIT catalog (Guo et al., 2013)
[d] IB-TFIT catalog (J. L. Donley et al. 2015, in preparation)
[e] GALEX DR6/7
[f] Measurements from two regions: GOODS-S and HUDF09. See the detail in Guo et al. (2013)
[g] Measurements from three regions: CANDELS wide, deep, and HUDF09. See Guo et al. (2013) for details.
[h] Measurements from two regions: 3D-HST and HUDF12. This is an updated version of Guo et al. (2013)

Table 2.2: Photometric Data in MUSYC+TENIS catalog

Filter	$\lambda_{\rm eff}$ (Å)	FWHM (Å)	5σ Limiting Depth (AB mag)	Instrument Telescope	Area
FUV[a]	1543	228	25.69	GALEX	2, 3
NUV[a]	2278	796	25.99	GALEX	2, 3
$U38$[b]	3706	357	25.33	WFI/ESO MPG	2, 3
U[b]	3528	625	25.86	ESO MPG/WFI	2, 3
B[b]	4554	915	26.45	ESO MPG/WFI	2, 3
V[b]	5343	900	26.27	ESO MPG/WFI	2, 3
R[b]	6411	1602	26.37	ESO MPG/WFI	2, 3
I[b]	8554	1504	24.30	ESO MPG/WFI	2, 3
z[b]	8989	1285	23.69	Blanco/Mosaic-II	2, 3
J[b]	12395	1620	22.44	Blanco/ISPI	2, 3
H[b]	16154	2950	22.46	ESO NTT/SofI	2, 3
K[b]	21142	3312	21.98	Blanco/ISPI	2, 3
J[c]	12481	1588	24.50	CFHT/WIRCam	2, 3
Ks[c]	21338	3270	23.90	CFHT/WIRCam	2, 3
3.6μm-SIMPLE[b,c]	35508	7432	23.89	Spitzer/IRAC	2, 3
4.5μm-SIMPLE[b,c]	44960	10097	23.75	Spitzer/IRAC	2, 3
5.8μm-SIMPLE[b,c]	57245	13912	22.42	Spitzer/IRAC	2, 3
8.0μm-SIMPLE[b,c]	78840	28312	22.50	Spitzer/IRAC	2, 3
IA427[b]	4253	210	25.01	Subaru	2, 3
IA445[b]	4445	204	25.18	Subaru	2, 3
IA464[b]	4631	216	24.38	Subaru	2, 3
IA484[b]	4843	230	26.22	Subaru	2, 3
IA505[b]	5059	234	25.29	Subaru	2, 3
IA527[b]	5256	243	26.18	Subaru	2, 3
IA550[b]	5492	276	25.45	Subaru	2, 3
IA574[b]	5760	276	25.16	Subaru	2, 3
IA598[b]	6003	297	26.05	Subaru	2, 3
IA624[b]	6227	300	25.91	Subaru	2, 3
IA651[b]	6491	324	26.14	Subaru	2, 3
IA679[b]	6778	339	26.02	Subaru	2, 3
IA709[b]	7070	321	24.52	Subaru	2, 3
IA738[b]	7356	324	25.93	Subaru	2, 3
IA768[b]	7676	366	24.92	Subaru	2, 3
IA797[b]	7962	354	24.69	Subaru	2, 3
IA827[b]	8243	339	23.60	Subaru	2, 3
IA856[b]	8562	324	24.41	Subaru	2, 3

[a] GALEX DR6/7
[b] MUSYC catalog (Cardamone et al., 2010b)
[c] TENIS catalog (Hsieh et al., 2012)

photometry in lower-resolution images. TFIT was also applied to the MIR data from the Spitzer Extended Deep Survey (SEDS; Ashby et al., 2013). This deblending yields more accurate photo-z and also increases the probability of making correct X-ray to IR associations (see Section 2.2). In addition, the Area 1 data include 18 IBs (Figure 2.3) at optical wavelengths provided by the MUSYC team[2] (Cardamone et al., 2010b). CANDELS collaborators (J. L. Donley et al. 2015, in preparation) have produced an IB-TFIT catalog with the same parameters used by G13. Despite being up to 2 magnitudes shallower than the rest of the optical data, the intermediate-band (IB) data are useful to identify emission lines, which can modify the choice of template best fitting the data and thus the photo-z (see Section 3.1.3). To these 36 bands we also added the near-UV (NUV) and far-UV (FUV) data from GALEX Data Release 6 and 7. The association between GALEX data and the WFC3/H-band catalog was done via positional matching within a radius of $1''$. About 5% of all sources and ∼25% of X-ray-detected sources have UV counterparts. The combined data, which we refer to as "$\mathrm{TFIT_{CANDELS+IB}}$" (Table 2.1), have 34930 sources with up to 38 bands for computing photo-z.

2. Areas 2+3: These areas differ in depth of X-ray coverage (Section 2.1.1) but have the same data sets otherwise. For the CDFS and ECDFS surrounding Area 1, we merged the following photometric catalogs via coordinate cross match, allowing a maximum separation of $1''$: (1) GALEX catalog (as above), (2) the original MUSYC catalog (Cardamone et al., 2010b), and (3) the J and K_s-band data from the Taiwan ECDFS Near-Infrared Survey[3] (TENIS; Hsieh et al., 2012). Although TENIS is no deeper than existing NIR data, the TENIS data are more homogeneous over the entire field and have slightly different transmission curves, increasing the wavelength coverage. The MIR data for Area 2 and 3 came from the Spitzer IRAC/MUSYC Public Legacy in ECDFS (SIMPLE; Damen et al., 2011). These data are shallower than the SEDS data available in Area 1. Table 2.2 lists the data sets used, and we refer to this dataset as "MUSYC+TENIS". There are 70049 sources in this photometry.

2.1.3 Spectroscopic Data

The availability of spec-z for a subgroup of sources is essential for computing reliable photo-z via SED fitting (Dahlen et al., 2013). A subset of spec-z can first be used for training under the assumption that they are representative of the entire population. A different subset can then be used for testing photo-z quality. For this work we cross matched the photometric catalogs to a compilation of spec-z (N. Hathi, private communication), allowing a maximum separation of $1''$. There are 2314 (∼7%) Area 1 sources that have reliable spec-z and 3880 (∼ 6%) such sources in Areas 2 and 3 (2016 in Area 2, 1864 in Area 3).

[2]Multi-wavelength Survey by Yale-Chile. The reduced images are available at `http://www.astro.yale.edu/MUSYC/`

[3]The TENIS data are available at `http://www.asiaa.sinica.edu.tw/~bchsieh/TENIS/About.html`

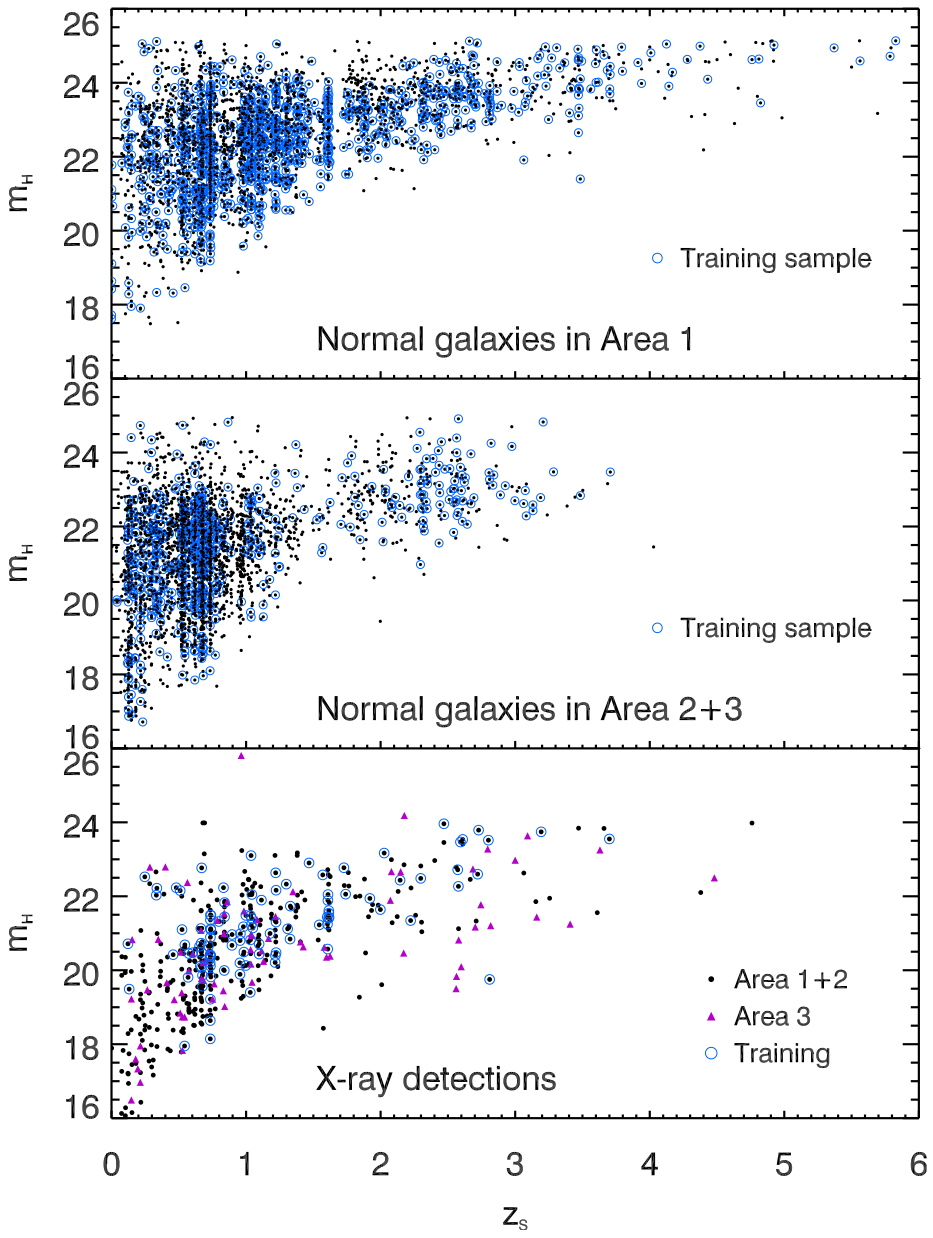


Figure 2.4: H-band magnitude as a function of spec-z for all objects with spectroscopic redshifts. Black dots in the top panel represent normal galaxies in Area 1, where $\mathrm{TFIT}_{\mathrm{CANDELS+IB}}$ data are available. The middle panel shows normal galaxies identified from the MUSYC catalog in Areas 2 and 3. Black dots in the bottom panel represent X-ray-detected sources in Areas 1 and 2, and magenta triangles denote sources detected in the shallower X-ray data in Area 3. Open blue circles in all three panels indicate objects used for training.

As discussed by Dahlen et al. (2013), optimal results are obtained when the templates used for the photo-z computation are calibrated on the photometry available for the spectroscopic training samples. For this reason the training samples should fully span the entire magnitude–redshift parameter space. Figure 2.4 shows that the 1000 sources randomly selected as our training samples are indeed spread over all redshift and magnitude ranges in the respective Areas. Because the photometry available in Area 1 differs from that in Area 2+3, two sets of training samples and computations of the zero-point offsets[4] were used.

For the X-ray sources, we forgo using the training sample for computing zero-point offsets and instead use it to sample the AGN population and help build the AGN-galaxy hybrid templates needed for proper SED fitting and photo-z measurement (Salvato et al., 2009). For this purpose, we randomly chose ∼25% of the 4Ms-CDFS detections with available spec-z over the entire range of redshift and magnitude that have CANDELS data and used them as the training set to build hybrid templates. The remaining ∼75% were used for unbiased testing of the results.

2.2 X-ray to optical/NIR/MIR Associations

X-ray source positions can differ between catalogs because of different methods adopted for data reduction and source detection. Our aim is not to judge which method of X-ray source detection is superior but rather to provide accurate photo-z for optical/NIR/MIR sources associated with X-ray sources. Associations between X-ray sources and possible counterparts were therefore done independently for each of the four X-ray catalogs (Sec 2.1.1), and duplicate sources were removed only at the end of the process as described below.

2.2.1 Comparing X-ray Catalogs

1. In Areas 1+2:

The major difference between the R13 and X11 catalogs is that R13 adopted a higher threshold for source detection. Despite that, there are some sources in the R13 catalog but not in X11. There are also astrometric differences, which can affect the association to an optical/NIR/MIR counterpart. Thus the redshift assigned to the X-ray source and also to the supposed counterparts can be different because different template libraries and priors were used for X-ray galaxies than for normal ones. In order to match X-ray catalogs, we shifted the X11 positions by $-0\farcs175$ in right ascension and $0\farcs284$ in declination[5] to register them to the optical frame (Giavalisco et al., 2004a). The R13 catalog is already on the

[4]Zero-point offset is the average for each photometric band of the difference between the photometry of training set objects and photometry predicted by the best-fit template at the object's redshift. The offset in each band depends on the set of templates used and the number of bands available.

[5]The original X11 positions are on the radio astrometric frame. The shifts needed to bring them to the optical frame are in Sec. 3.1 of the X11 paper.

Table 2.3: Catalogs used for redshift estimation and counterpart identification.

	Area 1	Area 2	Area 3
Cross matching	4Ms-CDFS-X11[e] 4Ms-CDFS-R13[f] CANDELS-TFIT MUSYC TENIS SEDS-IRAC[g]	4Ms-CDFS-X11 4Ms-CDFS-R13 MUSYC TENIS SIMPLE-IRAC[h]	250ks-ECDFS-L05[i] 250ks-ECDFS-V06[j] MUSYC TENIS SIMPLE-IRAC
Photo-z	CANDELS-TFIT[a] IB-TFIT[b] GALEX-UV	MUSYC[c] TENIS[d] GALEX-UV	MUSYC TENIS GALEX-UV
$N_{\rm spz}$[k]	2314	2016	1864

[a] Guo et al. (2013)
[b] J. L. Donley et al. (2015, in preparation)
[c] Cardamone et al. (2010b)
[d] Hsieh et al. (2012)
[e] Xue et al. (2011)
[f] Rangel et al. (2013)
[g] Ashby et al. (2013)
[h] Damen et al. (2011)
[i] Lehmer et al. (2005a)
[j] Virani et al. (2006)
[k] $N_{\rm spz}$ is the number of spec-z used in each Area (N. Hathi, private communication)

MUSYC optical frame and was not shifted. Figure 2.5a shows the coordinate differences before and after the astrometric shift.

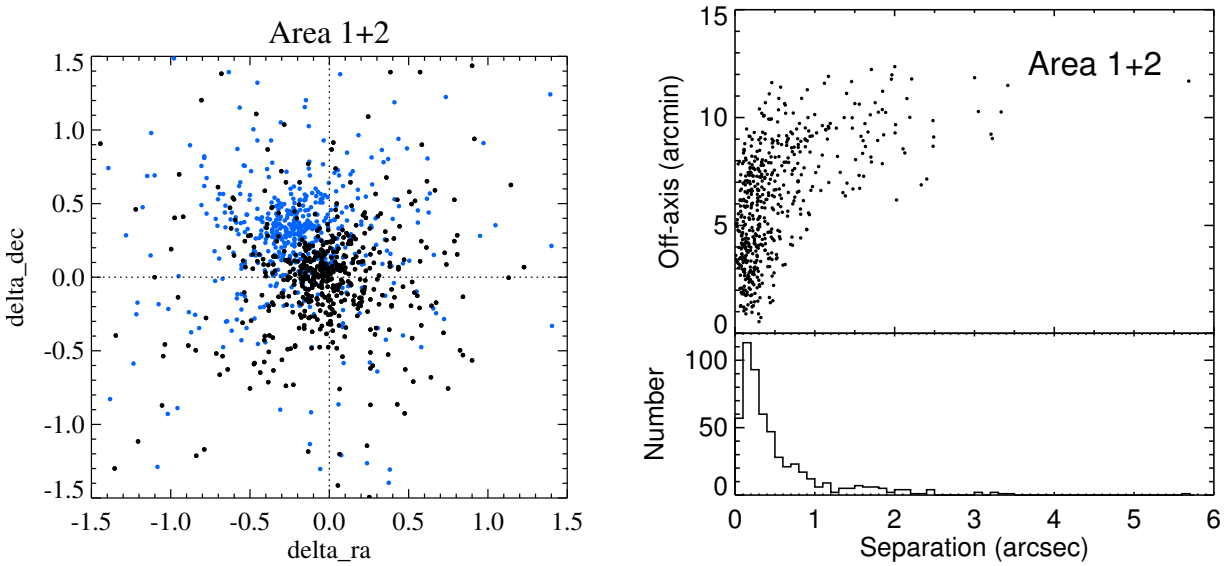
After astrometric shifting, we matched the X11 and R13 catalog coordinates, allowing a maximum distance of 10″. There are 545 sources in common with a maximum offset < 6″ as shown in Figure 2.5b. For these 545 sources, neither catalog has any additional X-ray source within 10″. As expected, all of the large offsets are for sources at large off-axis angles. For off-axis angles < 6′, the median coordinate offset is 0.″13, and except for one source, the maximum offset at any off-axis angle is < 3.″5. We treat each of the 545 matched sources as a single X-ray detection. 54% of these sources have a distance from each other larger than the positional error claimed for either of the catalogs. In addition, there are 195 sources detected by X11 but not R13 and 24 sources detected by R13 but not X11 for a total of 764 X-ray sources in Areas 1+2. As R13 mentioned, the unique sources to either of the two catalogs are mostly low-significance detections and therefore of lower reliability. In the following discussions, "X-" sources indicate those from X11 and "R-" those from R13.

2. In Area 3:

We adopted the Cardamone et al. (2008) astrometric calibration to align the V06 positions to the MUSYC and L05 catalogs, which were already in agreement. Figure 2.6a shows the coordinate differences before and after the astrometric shift. After the shift, the two catalogs have 366 source matches with offsets < 6″. These have a median separation of 0.″16 (Fig. 2.6b). We consider these 366 sources as the same X-ray detection. 12% of these sources have a separation that is larger than the positional error associated with either of the catalogs. In addition, there are 91 sources in the L05 catalog but not V06 and 38 sources in the V06 catalog but not L05 for a total of 495 X-ray sources in Area 3. A compilation of the four X-ray catalogs with their original positions and fluxes is available under [Surveys] > [CDFS] through the portal `http://www.mpe.mpg.de/XraySurveys`.

2.2.2 AGN counterparts determination method

We used a new association method based on Bayesian statistics which allows pairing of sources from more than two catalogs at once while also making use of priors. M. Salvato et al. (2015, in preparation) will give details, but in brief the code finds matches based on the equations developed by Budavári & Szalay (2008). Then additional probability terms based on the magnitude and color distributions are applied (see Naylor et al. 2013 for a similar approach). The code was developed in view of the launch of eROSITA (Merloni et al., 2012), where an expected million sources will be scattered over the entire sky and will have a non-negligible positional error and/or non-homogenous multi-wavelength coverage, conditions not optimal for association methods like Maximum Likelihood (e.g., Brusa et al. 2007; Luo et al. 2010; Civano et al. 2012). The new method provides the same quality of results as Maximum Likelihood in a much shorter time because matches are done simultaneously across all bands. Thus for example sources that are extremely faint or undetected in optical bands but brighter in the IRAC 3.6 μm band can be identified as counterparts in a single iteration.



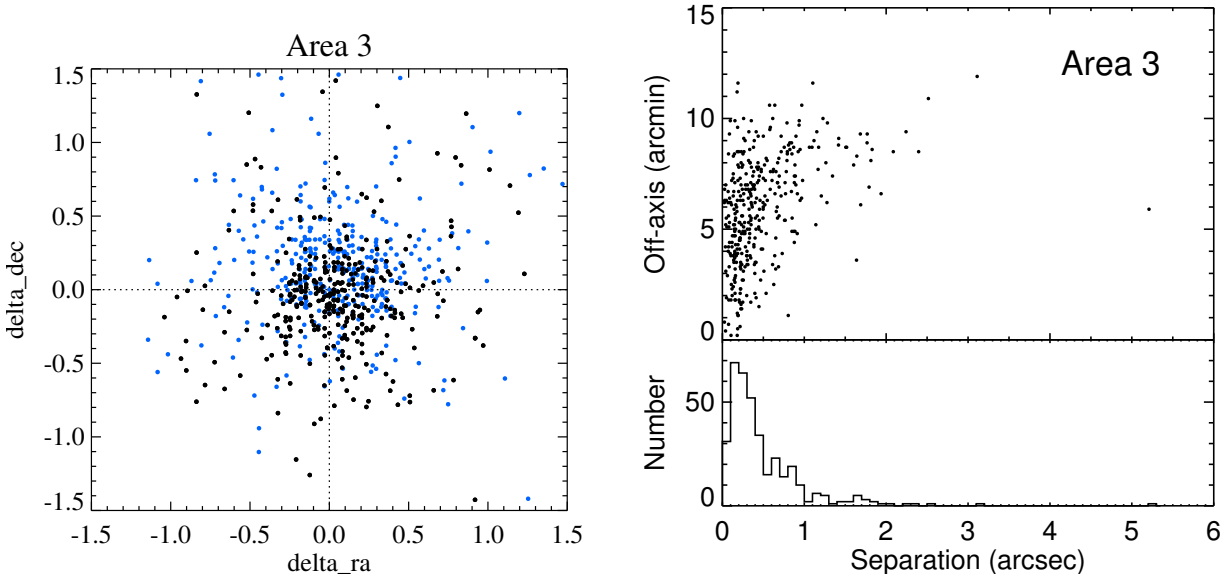
(a) (b)

Figure 2.5: Coordinate differences between X11 and R13 X-ray catalog. Left: Positional offset in R.A. and Dec. between R13 and X11 before (blue dots) and after (Black dots) the astrometric calibration. Right: The upper panel shows the off-axis angle from the *Chandra* aim point as a function of the angular offset for the 545 common sources in Area 1+2 in the two catalogs after the astrometric calibration. The lower panel shows the distribution of the angular separation.

For the 4Ms-CDFS sources (X11 and R13) located in Area 1, we used the CANDELS/H-selected catalog, TENIS/$J\&K_s$-selected catalog, MUSYC/BVR-selected catalog, and the deblended SEDS/IRAC 3.6 μm catalog. For the 4Ms-CDFS sources located in Area 2, we matched the X-ray sources to the TENIS/$J\&K_s$-selected catalog, MUSYC/BVR-selected catalog, and SIMPLE/IRAC 3.6μm catalog. The same set of these three catalogs was also used in Area 3 for finding the associations for the 250ks-ECDFS sources (L05 and V06). Table 2.3 summarizes the catalogs matched in each Area.

For each X-ray source (740 from X11, 569 from R13, 440 from L05, and 374 from V06), we considered all catalog objects lying within 4″ of the X-ray position and computed the posterior probability p that the given object is the correct counterpart. Figure 2.7 shows the distribution of the posteriors for all the possible associations in the three areas. In Area 1 where the data are deeper and better resolved, more than 98% of the X-ray sources have at least one association with $p > 0.7$, and we consider this p value the threshold for defining an association in all three areas. Area 3 has a distribution of p that reaches lower values (associations with $p < 0.7$), but because of the shallowness and lower resolution of the data, we consider to be unreliable. Our catalogs (see Sec. A) include the p value to allow users to define a stricter threshold, depending on the scientific use intended.

Figure 2.8 shows examples of ambiguous identifications. In all three cases shown,



(a) (b)

Figure 2.6: Coordinate differences between L05 and V06 X-ray catalogs. Left: Positional offset in R.A. and Dec. between L05 and V06 before (blue dots) and after (Black dots) the astrometric calibration. Right: The upper panel shows the off-axis angle from the *Chandra* aim point as a function of the angular offset for the 495 common sources in Area 3 in the two catalogs after the astrometric calibration. The lower panel shows the distribution of the angular separation.

a single X-ray source has two possible H-band associations with $p > 0.99$. Even the simultaneous use of deblended IRAC photometry from SEDS does not help in associating a unique counterpart. The example in the middle shows that despite the high resolution of the CANDELS images, the upper source is still blended, and probably a third component is present. If a further deblending were applied, the H flux would be split among multiple components, thus reducing the probability of the upper source being the right association. In practice, we attempted no further deblending and simply flagged these kinds of objects as sources with multiple counterparts. For these cases, in addition to the photo-z computed using normal galaxy templates, we also provide the values obtained by assuming that they are AGNs. The photo-z results reveal that $\sim 20\%$ of these close pairs have similar redshifts and may be associated with galaxy mergers or galaxy groups. However the majority of apparent pairs are projections of unrelated objects.

2.2.3 Results

Figure 2.9 shows the decision tree for X-ray source associations and computing photo-z, and Table 2.4 gives numbers for each case in each Area. There are four cases:

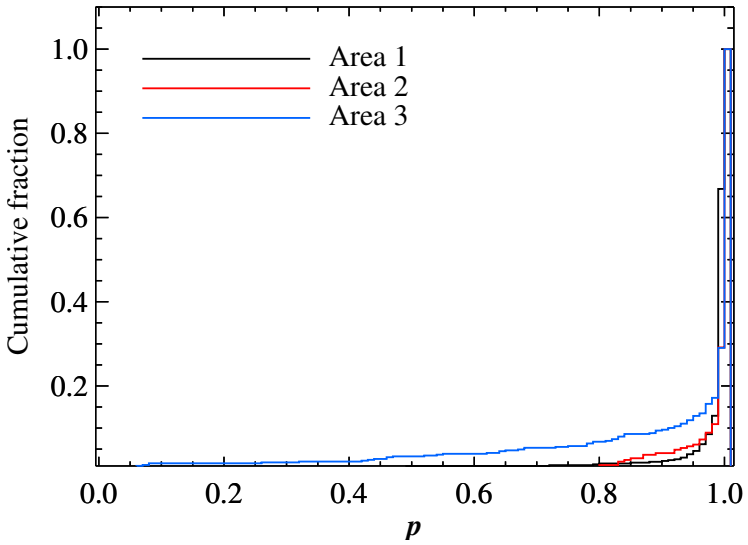


Figure 2.7: Cumulative fraction of the posterior p for the possible counterparts to the X-ray sources in Area 1, 2, and 3 as indicated in the legend.

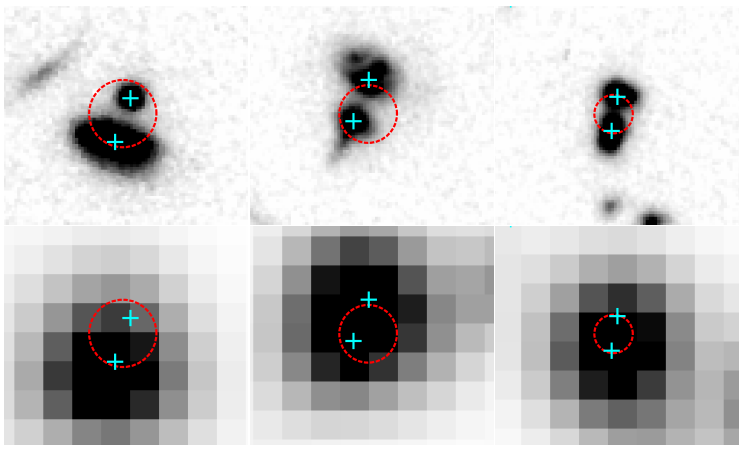


Figure 2.8: Three examples of multiple H-band associations (from left to right X-115, X-517, and X-224) in H-band (upper) and IRAC-3.6μm (lower). The size of each cutout is $5'' \times 5''$. The red circles are centered at the X-ray position with radius corresponding to the positional error. The cyan crosses indicate the positions of H-band detected sources from G13. These three cases have two H-band associations both with probabilities greater then 0.99. The uses of deblended IRAC photometry does not help in making a unique secure association.

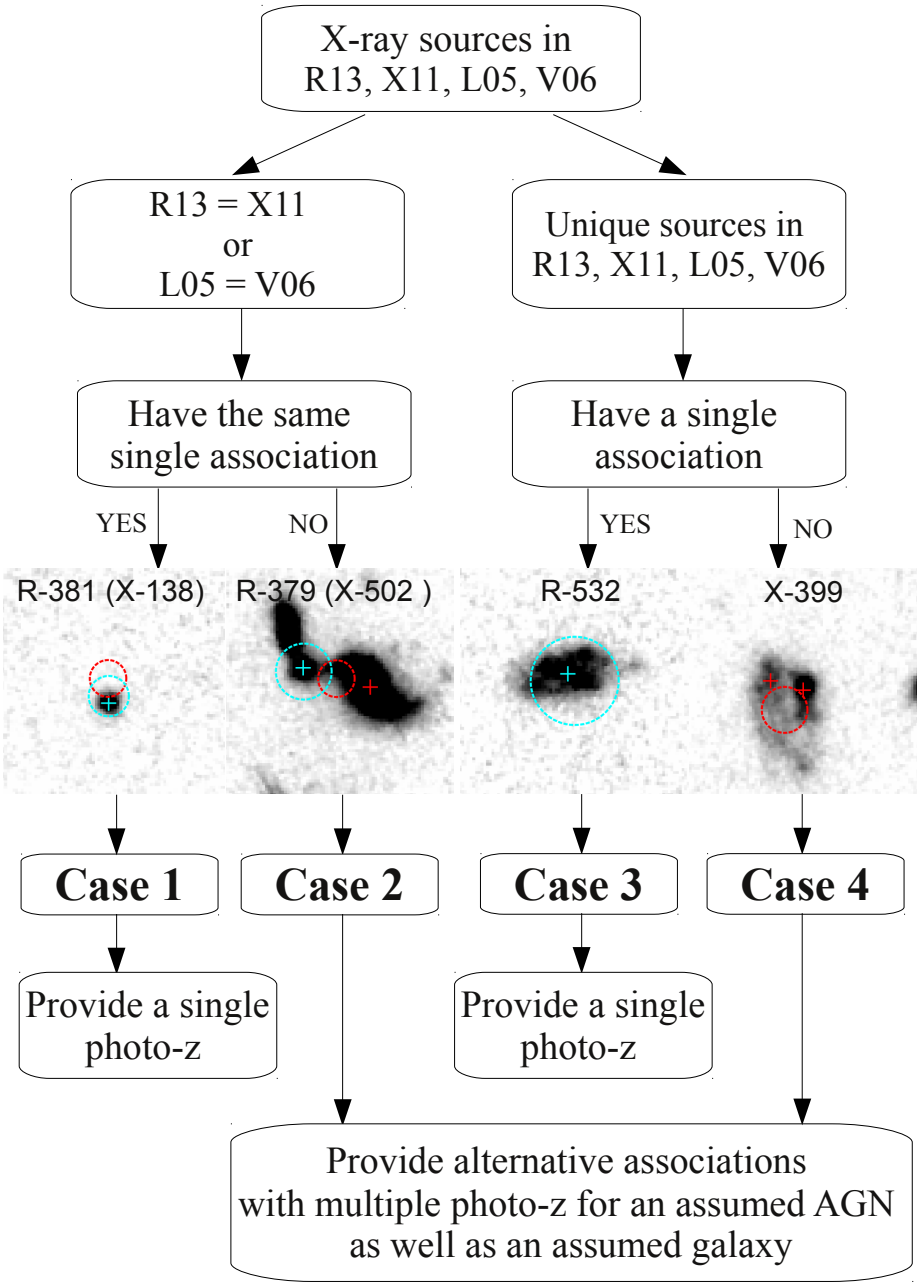


Figure 2.9: Flow chart of the process for four cases of X-ray to optical/NIR/MIR associations. H-band negative images ($5'' \times 5''$) are provided as examples for each case . Dashed-line circles show the X11 (red) and R13 (cyan) X-ray positions and positional uncertainties. Red and cyan crosses show the corresponding H-band counterparts.

- **Case 1:** An X-ray source in both catalogs with one optical/NIR/MIR association. Case 1 means the same unique association was chosen even though the X-ray catalog positions may differ between X11 and R13 or between L05 and V06. There are 714 of these sources in Areas 1+2+3.

- **Case 2:** An X-ray source in both catalogs with multiple optical/NIR/MIR associations. Case 2 can arise from two causes: (1) position differences in the X-ray catalogs may point to different counterparts, or (2) there may be more than one potential counterpart near the X-ray position(s), and we cannot tell which is the right one. Some of the latter may be blended sources with more than one galaxy contributing to the X-ray flux. In total, there are 181 case 2 sources in Areas 1+2+3. These sources are identified in the final catalogs, and counterpart photo-z are calculated using both AGN and normal galaxy SED templates (see Section A).

- **Case 3:** X-ray sources found in one catalog but not the other, having a unique counterpart. There are 235 of these sources in Areas 1+2+3.

- **Case 4:** X-ray sources found in only one catalog and having multiple possible counterparts. There are 77 such sources in Areas 1+2+3. As for Case 2, the catalogs identify all the possible counterparts and provide both AGN and normal galaxy photo-z results for each.

In summary, 1207 out of 1259 (∼96%) of the X-ray sources are associated with multi-wavelength counterparts, and 258 of them (∼21%) have multiple counterparts possible. There are 26 sources for which the counterpart is detected only in the IRAC bands, and no photo-z computation is possible for these because only two data points (3.6μm and 4.5μm) can be used. All the other sources have at least six photometric points, and a photo-z is provided. The photo-z catalog (see Sec. A) entry for each source indicates the number of photometric points used for the photo-z computation. The remaining 52 sources ($\sim$ 4%) either have no identifications in any of the optical/NIR/MIR catalogs ($\sim$ 1%) or have possible counterparts identified with $p < 0.7$ ($\sim$ 3%). For these sources, the photo-z are not available as well.

2.2.4 Comparison to previous results in Area 1+2

X11 used likelihood ratio matching to assign counterparts to 716 out of 740 X-ray sources in Areas 1+2. Our code and the newly available ancillary data give secure counterparts (with $p > 0.9$) for seven additional sources shown in Figure 2.10. Most of the new counterparts are offset from the X-ray position by 1–2 times the X-ray position uncertainty. The most likely reason for finding new identifications is having better imaging data available, but there remains a chance some of the X-ray sources are not real.

Figure 2.11 shows an example of a revised X-ray association. In this case low resolution catalogs give a single counterpart for the source (R-57=X-234) for either X-ray position.

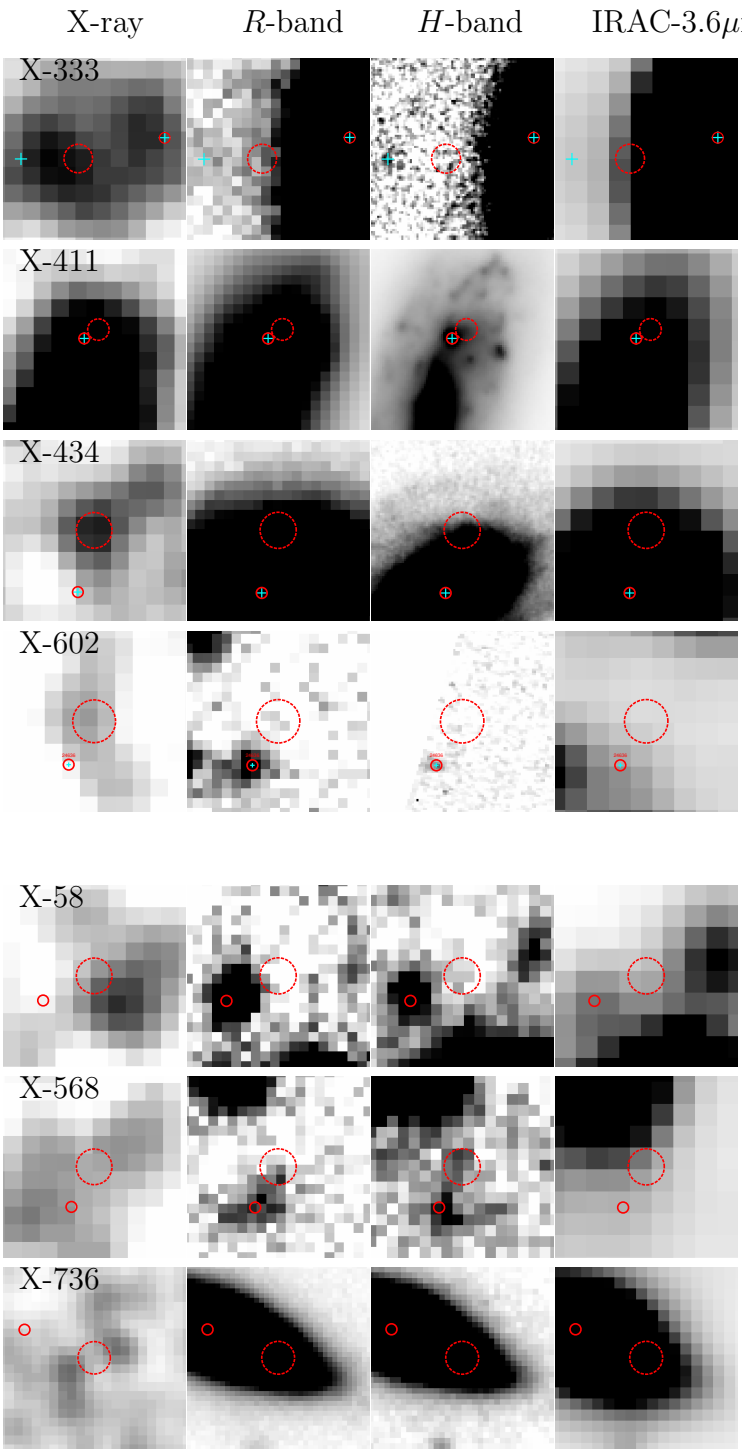


Figure 2.10: Multi-wavelength images of the seven sources from X11 for which we found new, secure ($p > 0.9$) counterparts. Wavelengths are indicated above each set of panels. The four sources in the upper group are in Area 1 and have CANDELS H-band images. The three sources in the lower group have no WFC3-H, and TENIS-K_s is shown instead. X-ray images are full-band from X11. The red dashed circles are centered at the X11 positions with their radii showing the corresponding positional uncertainty. Cyan crosses in the upper panels show all H-band detections, and the solid red circles show the catalog position of the chosen counterpart. All cutouts are $5'' \times 5''$ except that X-736 is $10'' \times 10''$.

Table 2.4: Results of X-ray to optical/NIR/MIR associations in ECDFS .

	N_x	Case1	Case2	Case3	Case4	N_{ctp}^{single}	N_{ctp}^{multi}	N_{ctp}	N_{ctp}^{multi}/N_{ctp}	N_{ctp}/N_x
Area 1	509	272	67	130	29	402	96	498	19%	98%
Area 2	255	170	29	35	12	205	41	246	17%	96%
Area 3	495	272	85	70	36	342	121	463	26%	94%
TOTAL	1259	714	181	235	77	949	258	1207	21%	96%

Note: N_x: Number of X-ray sources; N_{ctp}^{single}: Number of sources that have only one possible counterpart; N_{ctp}^{multi}: Number of sources that have more than one possible counterpart; N_{ctp}: Total number of sources for which at least a counterpart was found.

However, the high-resolution WFC3/H-band image reveals at least four sources close together, and the slightly different coordinates provided by X11 and R13 point to different but equally likely counterparts. This difference is mainly due to the catalogs chosen for cross matching rather than the matching method. The Bayesian method should in principle give the same result as the maximum likelihood method, but the ability to match several catalogs simultaneously greatly improves the efficiency of the matching.

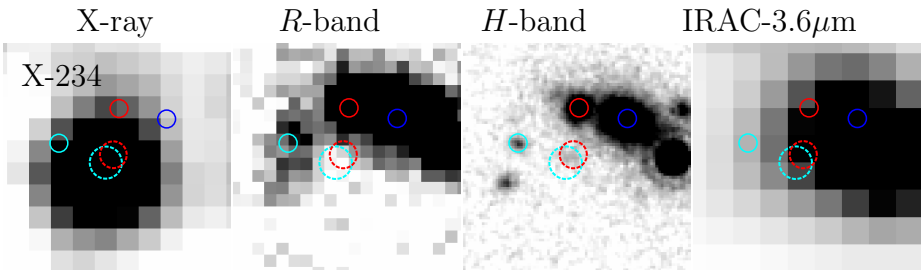
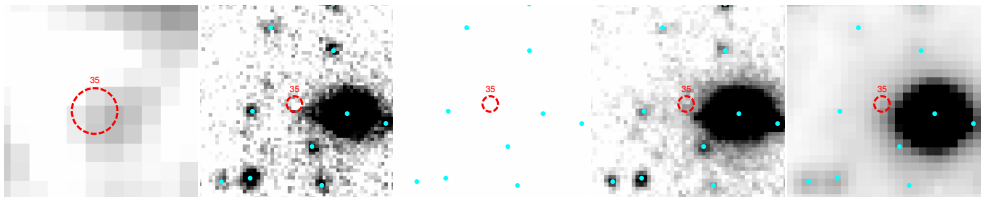


Figure 2.11: Negative images of the source R-57 (=X-234). Image wavelengths are indicated at the top, and each image is 5″ × 5″. Red dashed-line circles are centered at the position provided by X11 and cyan dashed-line circles at the position given by R13. Circle sizes indicate the respective X-ray position uncertainties. Red and cyan solid-line circles are the counterparts we assign to the two X-ray positions, and the blue circle indicates the counterpart assigned by X11.

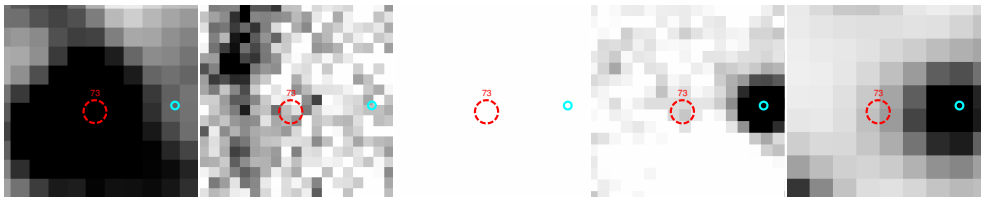
For the remaining 17 sources that are not identified (see Figure 2.12), eight of which were detected with X-ray false-positive threshold less than 1×10^{-6}. Referring to the Luo et al. (2010), one of the eight sources (i.e., X-73) is considered as the extreme X-ray/optical ratio source (EXO; Koekemoer et al., 2004) with $F_{0.5-8keV}/F_z \sim 50$. In general, most of the EXOs have clear detections in the IR and NIR, and are suggested be dusty galaxies at moderate redshifts ($z \sim 2-3$). Since there is no detection of the source X-73 in the deep IR and NIR, it is likely to be AGN at very high redshift ($z > 5$). The rest seven sources (X-35, 194, 477, 603, 643, 668, 688) at low-significance levels could be spurious

X-ray detections. If any of them is a real X-ray detection, then it could have a similar nature as source X-73 (Luo et al., 2010).

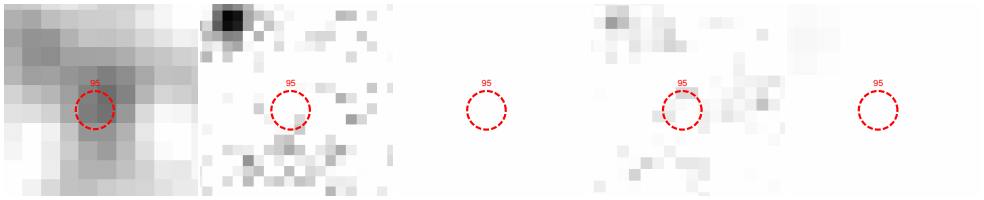
For the other nine X-ray sources at threshold greater than 1×10^{-6} (X-95, 98, 190, 280, 286, 366, 384, 408, 655), they are probably related to a star, galaxy groups and clusters, or off-nuclear sources (Luo et al., 2010; Xue et al., 2011). We visually inspected their optical, IR and X-ray images (as shown in Figure 2.12), and found that only source X-655 is likely to be an off-nuclear object. The rest of eight sources are probably galaxy groups or clusters as described in Luo et al. (2010) and Xue et al. (2011).



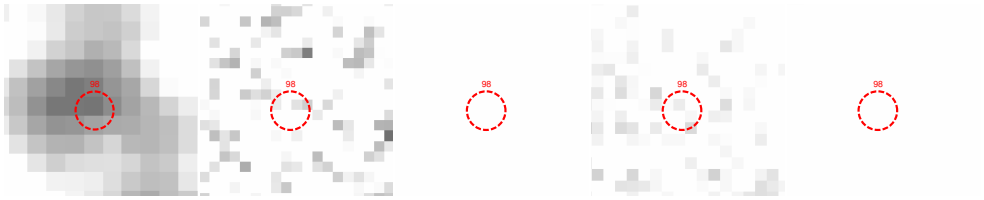
(a) X-35



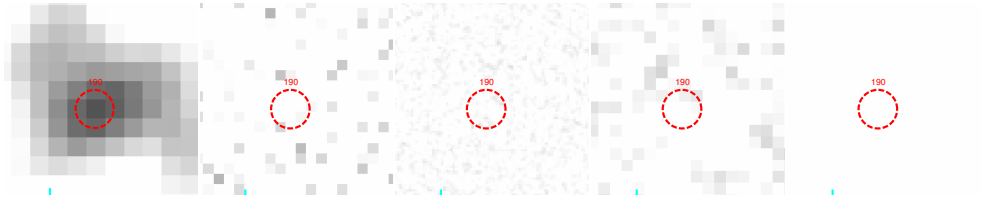
(b) X-73



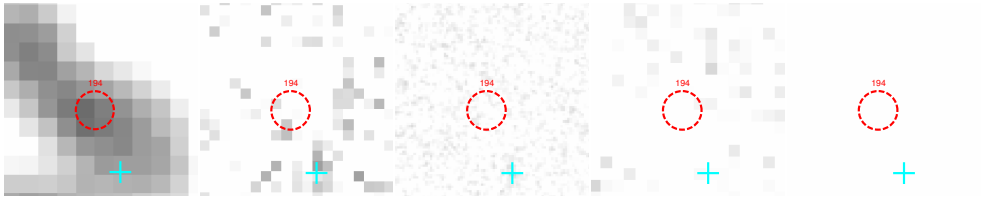
(c) X-95



(d) X-98

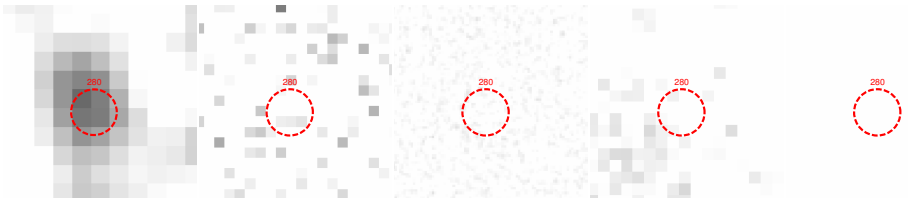


(e) X-190

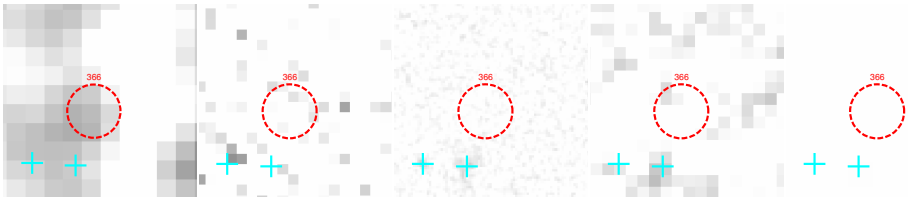


(f) X-194

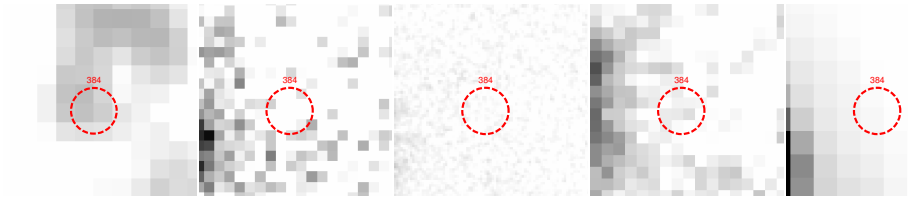
Figure 2.12: Cutouts of 17 unidentified X-ray sources from the X11 catalog. From left to right are the images in X-ray, R-band, H-band, Ks-band, and IRAC-3.6μm.



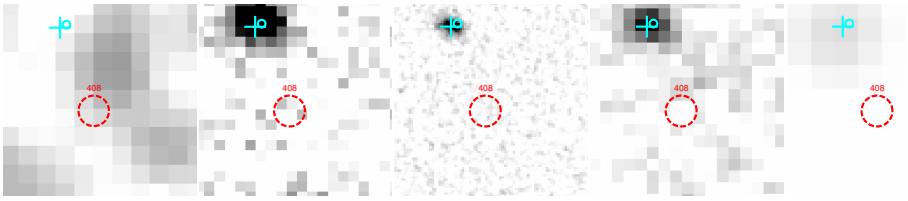
(g) X-280



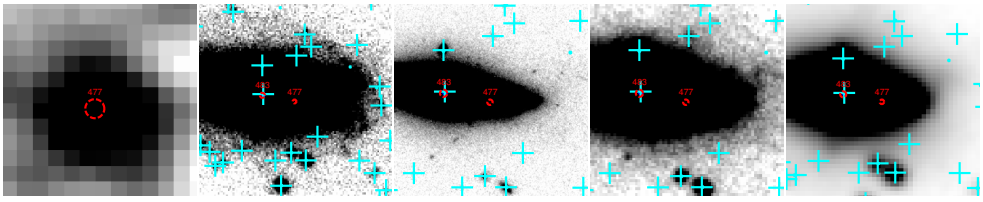
(h) X-366



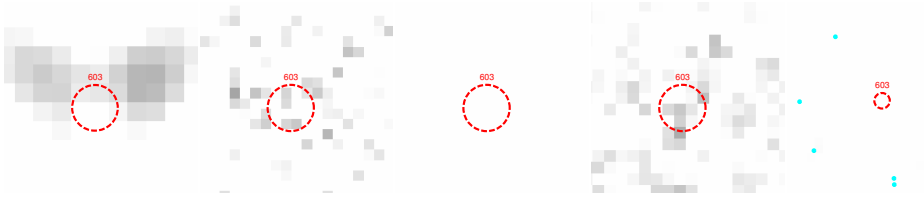
(i) X-384



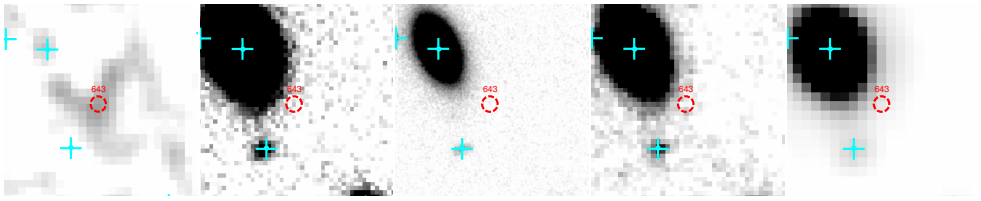
(j) X-408



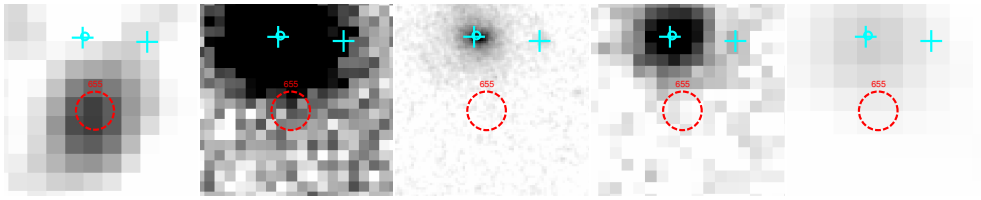
(k) X-477



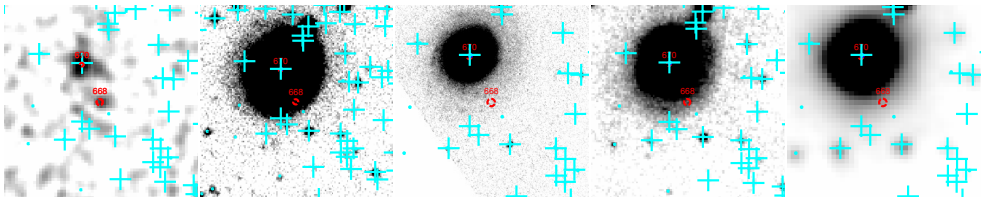
(l) X-603



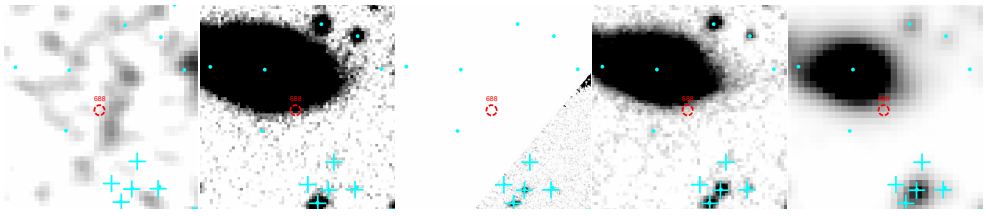
(m) X-643



(n) X-655



(o) X-668



(p) X-688

Chapter 3

Photometric redshift in the Chandra Deep Field South

Photo-z for galaxies and AGNs in the CDFS were computed using *LePhare* (Arnouts et al., 1999; Ilbert et al., 2006), which is based on a χ^2 template-fitting method (Arnouts et al., 1999, 2002). Applying three free parameters: redshift (z), template (T), and normalization factor (A), the function χ^2 is defined as

$$\chi^2(z,T,A) = \sum_{i=1}^{N} \left(\frac{F^i_{obs} - A \times F^i_{pred}(z,T))}{\sigma^i_{obs}} \right)^2 \tag{3.1}$$

where $F^i_{pred}(z,T)$ is the flux predicted for a template T at redshift z, F^i_{obs} is the observed flux, and σ^i_{obs} is the associated error for the considered filter i. Sum up N filters, the photo-z is obtained by the minimization of χ^2.

Depending on whether sources are the detected and non-detected in the X-ray surveys, four libraries of galaxy/AGN hybrids templates are utilized to estimate photo-z. (See the detailed descriptions for each library in the following Section 3.1 and Section 3.2). In addition to galaxy/AGN hybrids templates, we included a complete library of stellar templates as did following Ilbert et al. (2009) and Salvato et al. (2009). For each source, χ^2 is estimated for both galaxy/AGN hybrids and stellar templates. To assess the photo-z quality, we excluded possible stars, using the criterion that they are point-like and have $1.5 \times \chi^2_{\rm star} < \chi^2_{\rm agn/gal}$, where $\chi^2_{\rm star}$ and $\chi^2_{\rm agn/gal}$ are the minimum χ^2 obtained with stellar templates and galaxy/AGN hybrids, respectively.

Four extinction laws (those of Prevot et al. 1984, Calzetti et al. 2000, and two modifications of the latter, depending on the kind of templates) were used with $E(B-V)$ values of 0.00 to 0.50 in steps of 0.05 mag. Photo-z values were allowed to reach $z = 7$ (in steps of 0.01) as deep photometry allows us to reach high redshifts (see details given by Ilbert et al. 2009). The fitting procedure included a magnitude prior, forcing sources to have an absolute magnitude in rest B-band between -8 and -24. Photometric zero-point corrections were incorporated but never exceeded 0.1 mag. Note the we obtained zero-point offsets using normal galaxies photometry, but against this step for the AGNs because optical

variability can be intrinsic to the source and not accounted for in the photometry.

Quantifying the photo-z accuracy ($\sigma_{\rm NMAD}$), the percentage of the outliers (η), and the mean offset between photo-z and spec-z ($\rm bias_z$) was based on the spectroscopic samples with reliable quality (i.e., $\rm Q_{zs} < 2$, see the description for the value of $\rm Q_{zs}$ in Section A.4). The measure of $\sigma_{\rm NMAD}$ is the normalized median absolute deviation (NMAD):

$$\sigma_{\rm NMAD} = 1.48 \times \text{median}\left(\frac{|\Delta z|}{1+z_s}\right) \tag{3.2}$$

where z_s is spec-z, z_p is photo-z, and $\Delta z = (z_p - z_s)$. The value of $\sigma_{\rm NMAD}$ is calculated including outliers which are defined as:

$$\frac{|\Delta z|}{1+z_s} > 0.15 \tag{3.3}$$

Lastly, $\rm bias_z$ $(= \text{mean}(\frac{\Delta z}{1+z_s}))$ is calculated after excluding outliers.

3.1 Photo-z for Non X-ray detected Galaxies

This section focusses on the X-ray-undetected sources, which we refer to as "normal galaxies" even though some will in fact be AGNs.[1] The derived photo-z will be reliable to the extent of the sources exhibit normal galaxy SEDs at observed visible/infrared wavelengths. X-ray sources need more tuning for accurate photo-z and are discussed in Section 3.2 and Section 3.2.4.

3.1.1 Input catalogs and SEDs library

For the normal galaxies, we adopted the same templates, extinction laws, and absolute magnitude priors as Ilbert et al. (2009). The library includes 31 stellar population templates which were corrected for theoretical emission lines ([OII], [OIII], Hβ, Hα, and Lyα) by modeling the line fluxes at any redshift, template, and extinction. Based on Kennicutt (1998) calibration laws, as described in Ilbert et al. (2009), the [OII] emission line flux and UV luminosity (at 2300Å) are both related to the star formation rate (SFR) with the relations:

$$\text{SFR}(M_\odot \text{yr}^{-1}) = 1.4 \times 10^{-28} L_{\rm uv}(\text{erg s}^{-1}\ \text{Hz}^{-1}) \tag{3.4}$$

$$\text{SFR}(M_\odot \text{yr}^{-1}) = 1.4 \pm 0.4 \times 10^{-41} L_{\rm [OII]}(\text{erg s}^{-1}) \tag{3.5}$$

[1] A fraction of galaxies that host AGNs in their central regions may have X-ray emission below the detection threshold of our observations, but are identified at infrared and/or radio wavelengths or by emission line ratios (e.g., Hickox et al., 2009; Donley et al., 2012). For example in Hickox et al. (2009), with finite-sensitivity and limited-depth of observation in X-ray and IR, they found little overleapt between X-ray and radio AGNs, while there is $30\% - 50\%$ overlap between the X-ray and IR-selected AGNs. Note that the fraction of overlap could be changed with different observational depth.

Therefore [OII] line flux can be estimated by the translation:

$$\log(\mathrm{F}_{[\mathrm{OII}]}) = -0.4 \times M_{\mathrm{UV}} + 10.65 - \frac{\mathrm{DM(z)}}{2.5} \tag{3.6}$$

where $F_{[\mathrm{OII}]}$ is in units of 10^{-17} erg cm^{-2} s^{-1}, DM is the distance modulus, and M_{UV} is the dust-corrected UV (2300Å) absolute magnitude.

With the predicted [OII] flux from the procedure above at every redshift, template, and extinction, we can estimate the other line fluxes from the ratios of [OIII]/[OII]=0.36, Hβ/[OII]=0.61, Hα/[OII]=1.77, and Lyα/[OII]=2 (McCall et al., 1985; Kennicutt, 1998; Mouhcine et al., 2005; Moustakas et al., 2006; Ilbert et al., 2009).

All the normal galaxies were selected from either the CANDELS-TFIT catalog or the MUSYC catalog (Sec. 2.1.2). Photo-z are based on $\mathrm{TFIT}_{\mathrm{CANDELS+IB}}$ photometry for sources detected in the CANDELS-TFIT catalog and otherwise on MUSYC+TENIS photometry. The majority of normal galaxies have $\mathrm{TFIT}_{\mathrm{CANDELS+IB}}$ photometry in Area 1 but only MUSYC+TENIS photometry in Areas 2+3. Table 3.1 gives these measures of photo-z quality for the global samples and for subsamples split into magnitude and redshift bins.

3.1.2 Results

• Area 1

In Area 1, the overall outlier fraction of ∼4.1% in this region is comparable to the most recent work by the CANDELS team (Dahlen et al., 2013). However, Table 3.2 shows that adding deblended IB photometry (i.e., $\mathrm{TFIT}_{\mathrm{CANDELS+IB}}$) improves the accuracy to $\sigma_{\mathrm{NMAD}} = 0.011$ (from $\sigma_{\mathrm{NMAD}} = 0.026$ by Dahlen et al. 2013) and $\mathrm{bias}_z = -0.000$ (from $\mathrm{bias}_z = -0.005$). Dahlen et al. (2013) used the CANDELS-TFIT data (no IB data is included) and obtained a better result than using only the ground-based data. However, their result is not as good as the photo-z quality we have achieved using the combined data sets (i.e., $\mathrm{TFIT}_{\mathrm{CANDELS+IB}}$). Adding the IB data improves our results not only in accuracy but also in outlier fraction, even for the fainter subset of the sample. Figure 3.1 illustrates the overall results in Area 1.

The outlier fractions and scatter are larger for the fainter galaxies , but the bias is only a weak function of source magnitude (Table 3.1). The bias is, however, larger for $z > 1.5$ galaxies than for those at lower redshifts. The scatter and outlier fraction are also larger at $z > 1.5$, but this mostly reflects the typically fainter magnitudes of the more distant sources.

• Areas 2 and 3

Outside the CANDELS area, the photo-z quality using MUSYC+TENIS photometry is similar to that of Cardamone et al. (2010b). Figure 3.2 illustrates the results. The brighter and lower-redshift subsets have photo-z quality almost as good as in Area 1 (see Table 3.1), but fainter galaxies have a higher outlier fraction. This is just as expected from

Table 3.1: Photo-z Quality for Normal Galaxies in Area 1+2+3

	Area 1				Area 2+3				Area 1+2+3			
	N	bias_z	σ_{NMAD}	$\eta(\%)$	N	bias_z	σ_{NMAD}	$\eta(\%)$	N	$bias_z$	σ_{NMAD}	$\eta(\%)$
Total	1995	-0.001	0.012	4.06	3587	0.002	0.010	4.88	5582	0.001	0.010	4.59
$R < 23$	583	0.003	0.008	1.37	2474	0.001	0.008	1.82	3057	0.002	0.008	1.73
$R > 23$	1412	-0.003	0.015	5.17	1113	0.003	0.013	11.68	2525	-0.001	0.014	8.04
$H < 23$	1333	-0.000	0.011	3.15	2457	0.002	0.009	3.05	3790	0.001	0.009	3.09
$H > 23$	662	-0.002	0.016	5.89	1130	0.001	0.011	8.85	1792	0.000	0.013	7.76
$z < 1.5$	1659	0.001	0.011	3.62	3457	0.002	0.009	4.57	5116	0.002	0.010	4.26
$z > 1.5$	336	-0.013	0.022	6.25	130	-0.007	0.030	13.08	466	-0.011	0.025	8.15

Table 3.2: Comparison of Photo-z Results for Normal Galaxies in Area 1

		$\text{TFIT}_{\text{CANDELS+IB}}$			MUSYC+TENIS			Cardamone+2010			Dahlen+2013		
	N	bias_z	σ_{NMAD}	$\eta(\%)$	bias_z	σ_{NMAD}	$\eta(\%)$	bias_z	σ_{NMAD}	$\eta(\%)$	bias_z	σ_{NMAD}	$\eta(\%)$
Total	1541	0.000	0.011	2.14	0.003	0.012	3.96	0.000	0.011	3.96	-0.005	0.026	2.47
$R < 23$	506	0.003	0.009	0.79	0.002	0.008	0.79	0.002	0.008	0.99	-0.002	0.026	0.99
$R > 23$	1035	-0.002	0.013	2.80	0.003	0.016	5.51	-0.001	0.016	5.41	-0.006	0.026	3.19
$H < 23$	1064	0.001	0.010	1.60	0.003	0.010	2.07	0.000	0.010	2.07	-0.006	0.027	1.97
$H > 23$	477	-0.002	0.014	3.35	0.002	0.021	8.18	0.000	0.022	8.18	-0.001	0.024	3.56
$z < 1.5$	1308	0.002	0.010	2.14	0.004	0.011	3.13	0.002	0.010	2.98	-0.005	0.026	2.45
$z > 1.5$	233	-0.014	0.019	2.15	-0.002	0.030	8.58	-0.008	0.045	9.44	-0.002	0.023	2.58

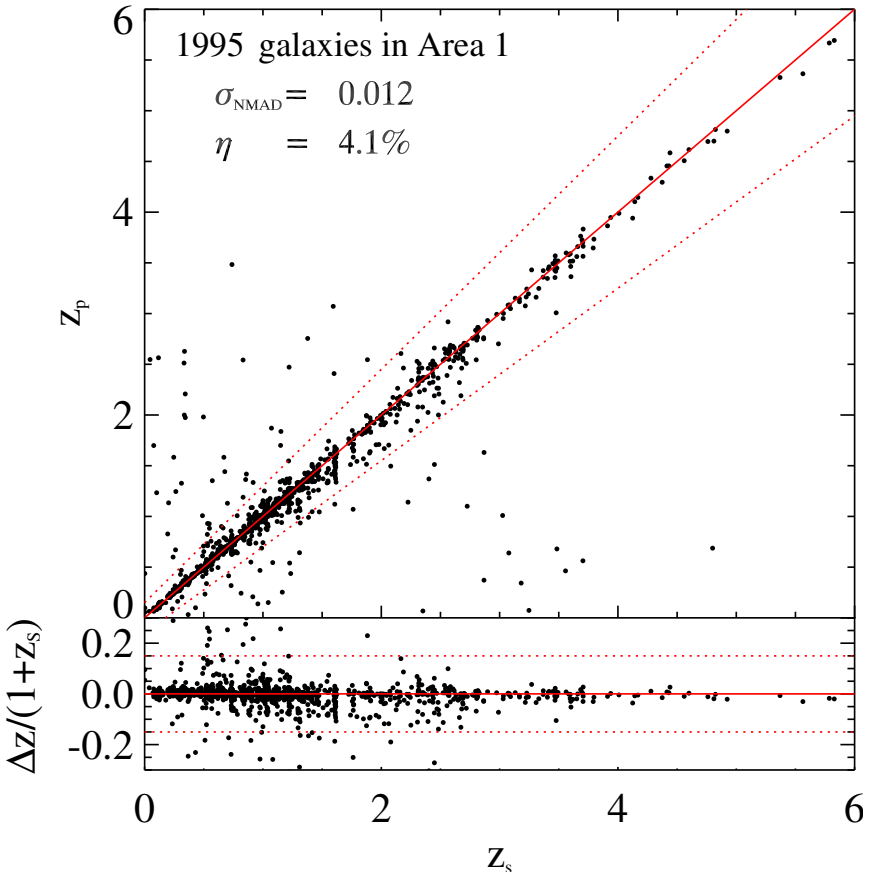


Figure 3.1: Upper panel: Photo-z vs spec-z for normal galaxies (i.e., non X-ray detections) in Area 1. Dots represent all normal galaxies with spec-z in Area 1. The solid line represents $z_p = z_s$; the dotted lines represent $z_p = z_s \pm 0.15(1 + z_s)$. Lower panel: Same but plotted as the difference $\Delta z \equiv (z_p - z_s)$.

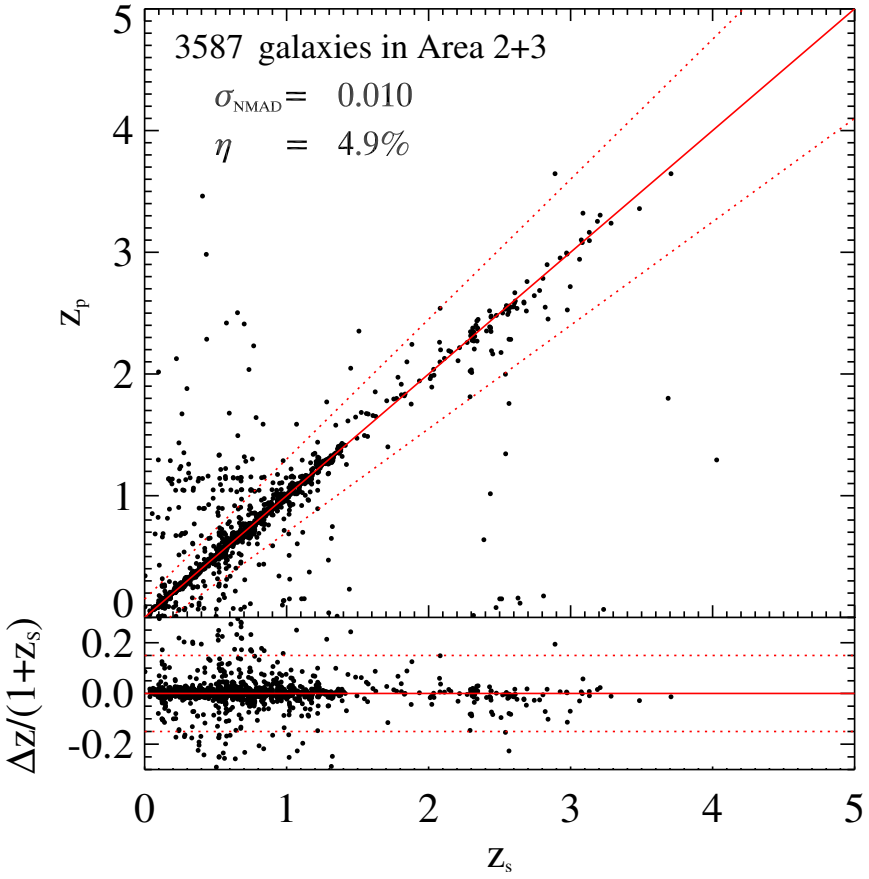


Figure 3.2: Photo-z vs spec-z for normal galaxies in Areas 2+3.

.

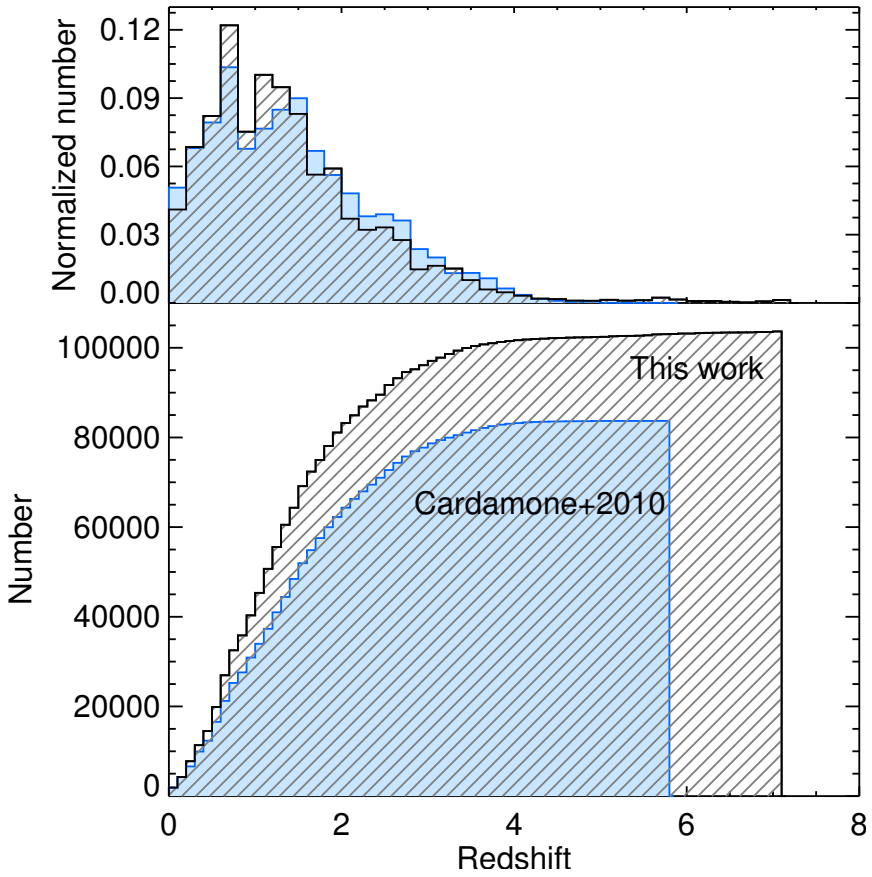


Figure 3.3: Upper panel: normalized photo-z distribution for normal galaxies. The grey hatched area shows results of this work, and the blue shaded area shows results of Cardamone et al. (2010b). Lower panel: cumulative number of normal galaxy the photo-z redshifts for this work and for Cardamone et al. (2010b) as labeled.

the tests in Section 3.1.2.

• Global result

The entire ECDFS (Areas 1+2+3) contains $\sim$ 104000 normal galaxies that have photo-z up to $z \sim 7$. Globally we reached the photo-z accuracy of 0.010 with an outlier fraction of 4.6% . Figure 3.3 shows the advantage of using deep WFC3 NIR to detect more sources in total and especially at $z > 2$. An interesting paradox is that we actually have a slightly lower fraction of sources at $z > 1.5$ than Cardamone et al. (2010b). This is probably because their higher outlier fraction, lacking deep NIR data, leads to more outliers with apparent $z > 1.5$.

3.1.3 Discussion

In the discussion below, we use the large spectroscopic sample of normal galaxies to demonstrate the importance of using deep WFC3/NIR data, intermediate-band data, and emission lines in the templates. These results can be applied also to the X-ray detected sources in the following sections.

• Impact of HST/WFC3 deep NIR data

The enormously decreased outlier fraction in Area 1 is due to the use of deep HST/WFC3 data and the deblended IB photometry. Table 3.2 gives data quality measures for 1541 sources in common using various data sets. Using only MUSYC+TENIS, but not the deep $\mathrm{TFIT}_{\mathrm{CANDELS+IB}}$ data, produces the same data quality as Cardamone et al. (2010b) as expected. However using the $\mathrm{TFIT}_{\mathrm{CANDELS+IB}}$ photometry decreases the outlier fraction from $\sim$4% to $\sim$2%, and the decrease is most substantial (more than a factor of two) for the faint and distant sources (declining from $\sim$9% to $\sim$2%). Figure 3.4 shows the comparison and in particular the decrease in outliers at $z_s > 2$. The difference mainly comes from the use of deep space-based data (i.e., CANDELS). The NIR data from the CANDELS catalog is deeper than the other two ground-based catalogs, and the related errors are much smaller. As shown in Figure 3.5, most of outliers from Cardamone et al. (2010b) in Figure 3.4 have large errors in J-band from the MUSYC photometry. This can cause degeneracy while running the fitting and increase the outlier number. Therefore, with the WFC3 data, the smaller error helps us to decrease the degeneracy and greatly improve our photo-z result. However, for those outliers both in our work and in Cardamone et al. (2010b) (i.e., blue circles in Figure 3.4 and black crosses in Figure 3.5), using deep CANDELS data didn't help to improve the photo-z results. In these cases, observational depth and the related flux error can not explain the inconsistency between photo-z and spec-z. This inconsistency could be the result of mismatching between the photometric data and spectroscopic sample in both of the works.

• The Impact of Intermediate-band Photometry

Previous work has shown the importance of IB photometry for photo-z, especially because IB data can show the presence or absence of emission lines in galaxy SEDs. For example Ilbert et al. (2009) showed that including IBs improved photo-z accuracy from 0.03 to 0.007 for normal galaxies with $i^+ < 22.5$ in the COSMOS field. Cardamone et al. (2010b) found similar results in the ECDFS. For AGNs, Salvato et al. (2009) showed that for both extended and point-like X-ray sources in COSMOS, accuracies and outlier fractions were substantially better when IBs were included.

In the CDFS, the IB photometry is much shallower than the NIR data from CANDELS (Table 2.1). To examine whether the shallow IB data are helpful or not in this case, we recomputed the Area 1 photo-z with exactly the same CANDELS-TFIT dataset (Guo et al., 2013) used by Dahlen et al. (2013), i.e., without IBs. Results of the comparisons are given in Table 3.2 and Figure 3.6.

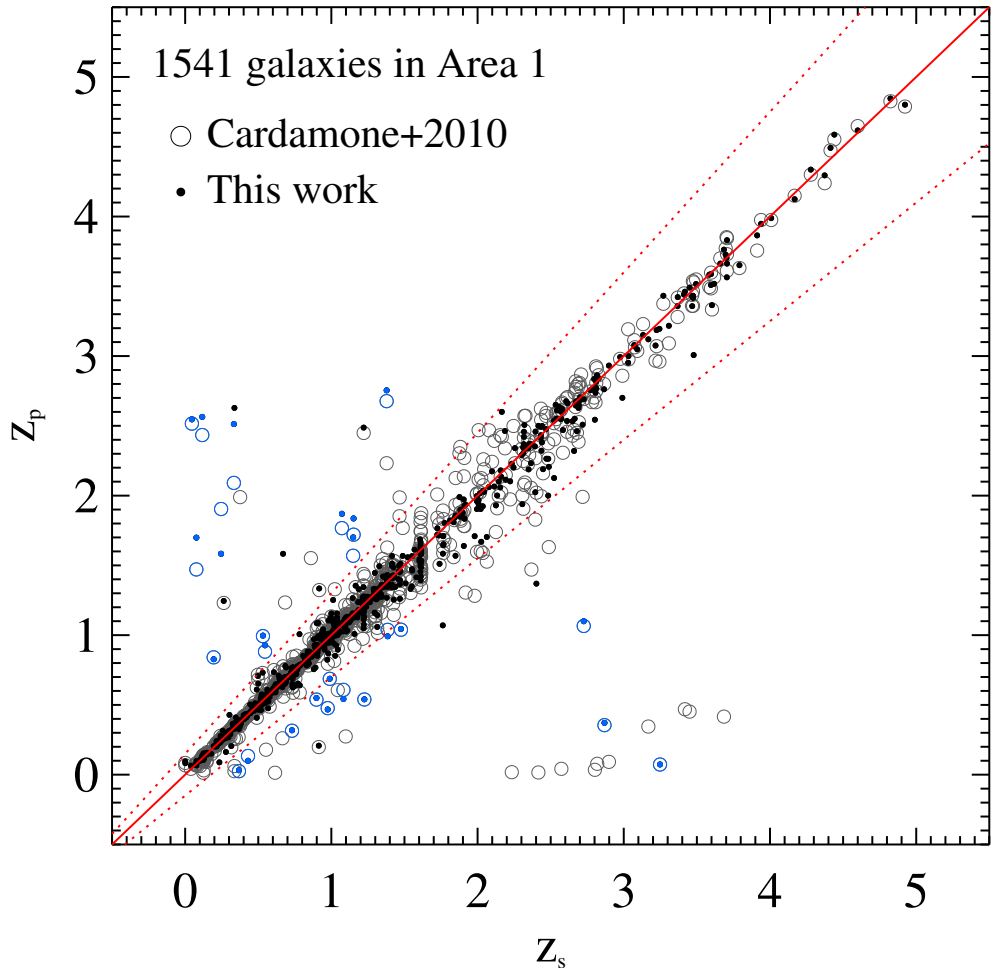


Figure 3.4: Photo-z vs. spec-z for 1541 common galaxies between this work and Cardamone et al. (2010b) in Area 1. Black dots are results from this work, and grey open circles are results from Cardamone et al. (2010b). Blue dots and blue open circles indicate objects that outliers both in our work and in that of Cardamone et al. (2010b).

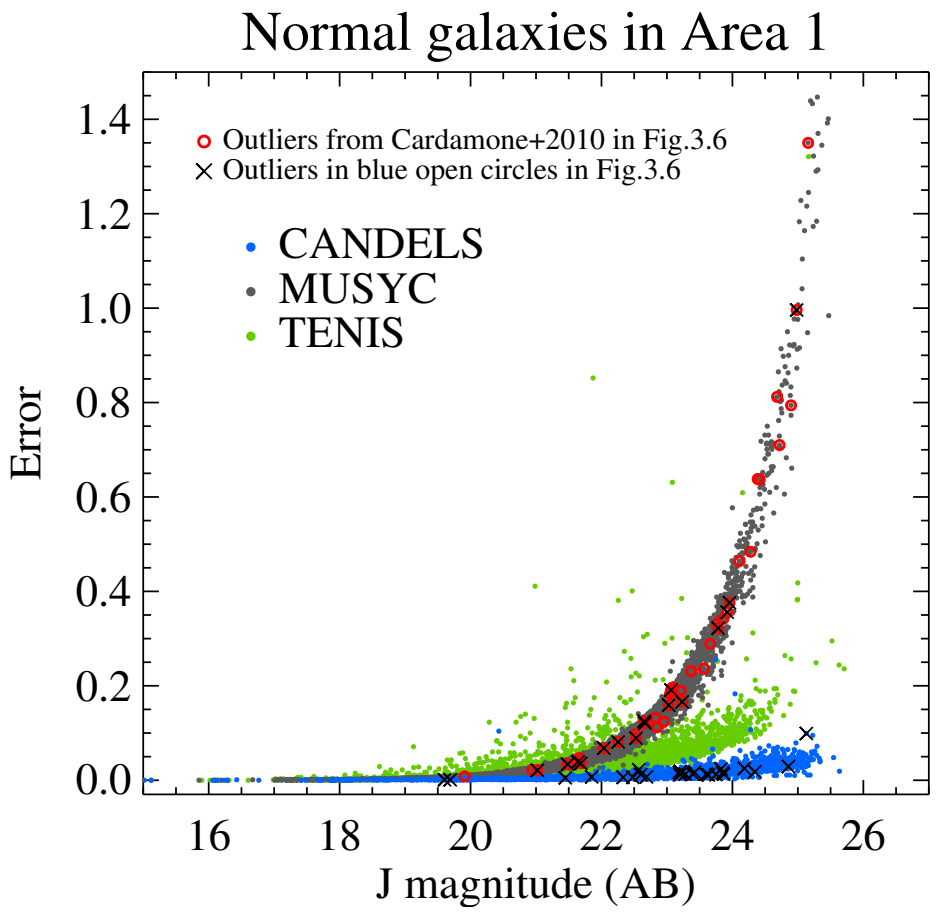


Figure 3.5: Photometric errors vs. J-band magnitude in three catalogs: MUSYC (grey), TENIS (black), and CANDELS (blue). The red open circles are the outliers from Cardamone et al. (2010b) in Figure 3.4; the black crosses are the outliers that both in our work and in Cardamone et al. (2010b) (blue open circles in Figure 3.4).

Without the IBs, the outlier fraction is 5%, accuracy is 0.037, and $\mathrm{bias}_z = -0.010$. These are similar to the results of Dahlen et al. (2013) "method 11H," which used the same code and library as in this thesis. The negative value of bias_z indicates underestimation of photo-z on average. This results in lower galaxy luminosities and incorrect rest-frame colors. As discussed by Rosario et al. (2013), these may lead to incorrect measurements of galaxy ages and stellar populations. The use of IB data improve the accuracy and mean offset substantially, creating a narrower and more symmetric peak of photo-z values around the spec-z (Fig. 3.6).

Intermediate bands should be most important for objects that have strong emission lines in their spectra. Strong emission lines can arise either from vigorous star formation or an AGN. To quantify the effect, we applied (inverse) BzK selection (Daddi et al., 2004) to define a sample of star-forming objects among those with reliable spec-z. In order to extend the selection at high redshift, we applied the revised BzK criterion as defined by

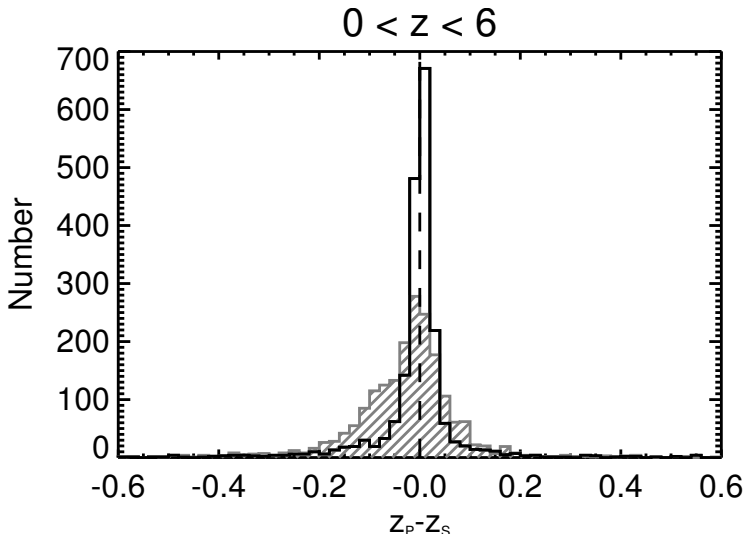


Figure 3.6: Distribution of $(z_{\rm p} - z_{\rm s})$ with and without IB data. The histograms show $z_p - z_s$ distribution for all galaxies with spec-z in Area 1. The black line shows results for photo-z computed with IB photometry (with $rmbias_z = -0.001$), and the hatched area shows results for the same galaxies without using IB data (with $rmbias_z = -0.010$).

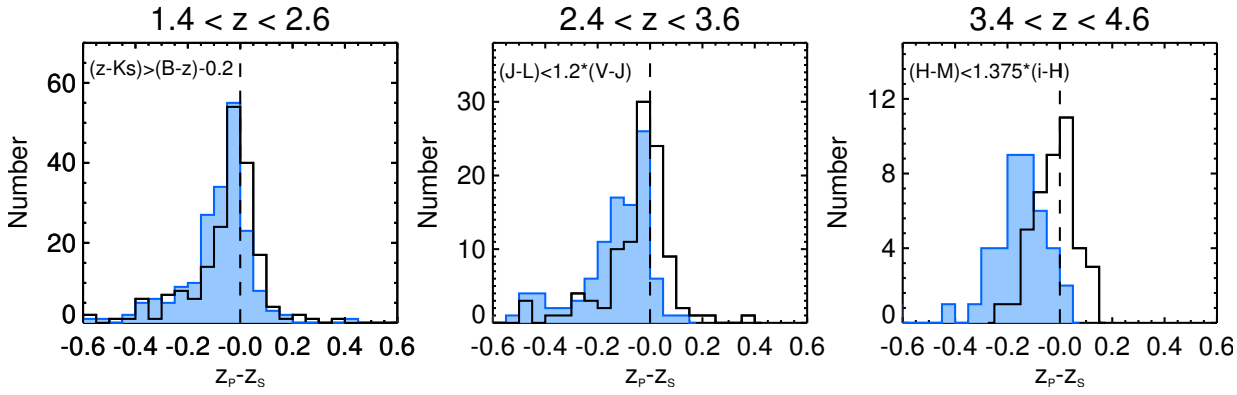


Figure 3.7: Distribution of $(z_{\rm p} - z_{\rm s})$ for Bzk-selected star-forming galaxies with and without IB data. The histograms show $(z_{\rm p} - z_{\rm s})$ distributions in various redshift bins as indicated above each panel. The Black lines show the distributions for photo-z computed with IB photometry, and blue areas show distributions for the same galaxies without using IB photometry. All data are from Area 1.

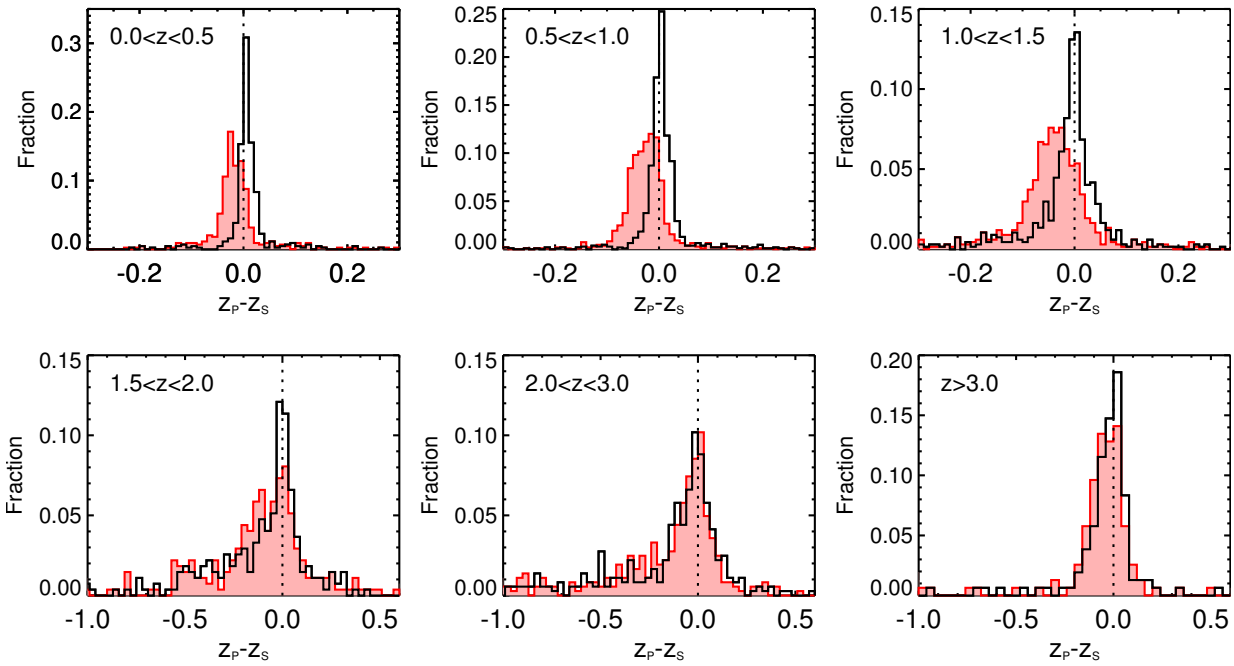


Figure 3.8: $(z_\mathrm{p} - z_\mathrm{s})$ distribution in various redshift bins for normal galaxies in Area 1. The photo-z are computed using the $\mathrm{TFIT}_{\mathrm{CANDELS+IB}}$ photometric catalog and using the templates with (black solid line) and without (red solid lines) emission line contributions. The upper panel shows that the emission lines are useful particularly at $z < 1.5$.

Guo et al. (2013).[2] Fig. 3.7 shows the resulting distributions of photo-z minus spec-z. At all redshifts, the distribution including IB is more peaked and symmetric around zero when IBs are included.

• Impact of Emission Lines in the Templates

Ilbert et al. (2009) demonstrated the importance of taking emission lines into account for photo-z. Including lines in the templates improved photo-z accuracy by a factor of 2.5 for bright ($i^+ < 22.5$) galaxies in the COSMOS field. The same effect is seen in the deeper $\mathrm{TFIT}_{\mathrm{CANDELS+IB}}$ data as shown in Figure 3.8. Although outlier numbers remain similar ($\sim$4%) whether emission lines are included in the templates or not, the distributions of $(z_\mathrm{p} - z_\mathrm{s})$ change. At $z < 1.5$, including emission lines gives much narrower peaks and lower bias. At $z > 1.5$, the improvement is less than at lower redshifts. Possible reasons are: (a) the contribution of the emission lines are diluted when observed in the NIR bands, which have broader bandwidths than optical bands; (b) the recipes for adding emission lines to the templates may be wrong for high-redshift galaxies; and/or (c) the IB data may be too

[2]The exact criteria were (1) $(z - K_s) > (B - z) - 0.2$ in the redshift range $1.4 < z < 2.6$; (2) $(J - L) > 1.2 \times (V - J)$ in the redshift range $2.4 < z < 3.6$; (3) $(H - M) > 1.375 \times (i - H)$ in the redshift range $3.4 < z < 4.6$. Symbols B, V, i, z, J, H, K_s, L, M refer to F435W, F606W, F775W, F850LP, F125W, F160W, ISAAC K_s, IRAC 3.6 μm, IRAC 4.5 μm, respectively.

shallow to affect the high-redshift (and therefore faint) sources. However, even at $z > 1.5$, the photo-z accuracy still shows a factor of 1.5 improvement ($\sigma_{\rm NMAD}$ decreasing from 0.032 to 0.021) when emission lines are included in the templates.

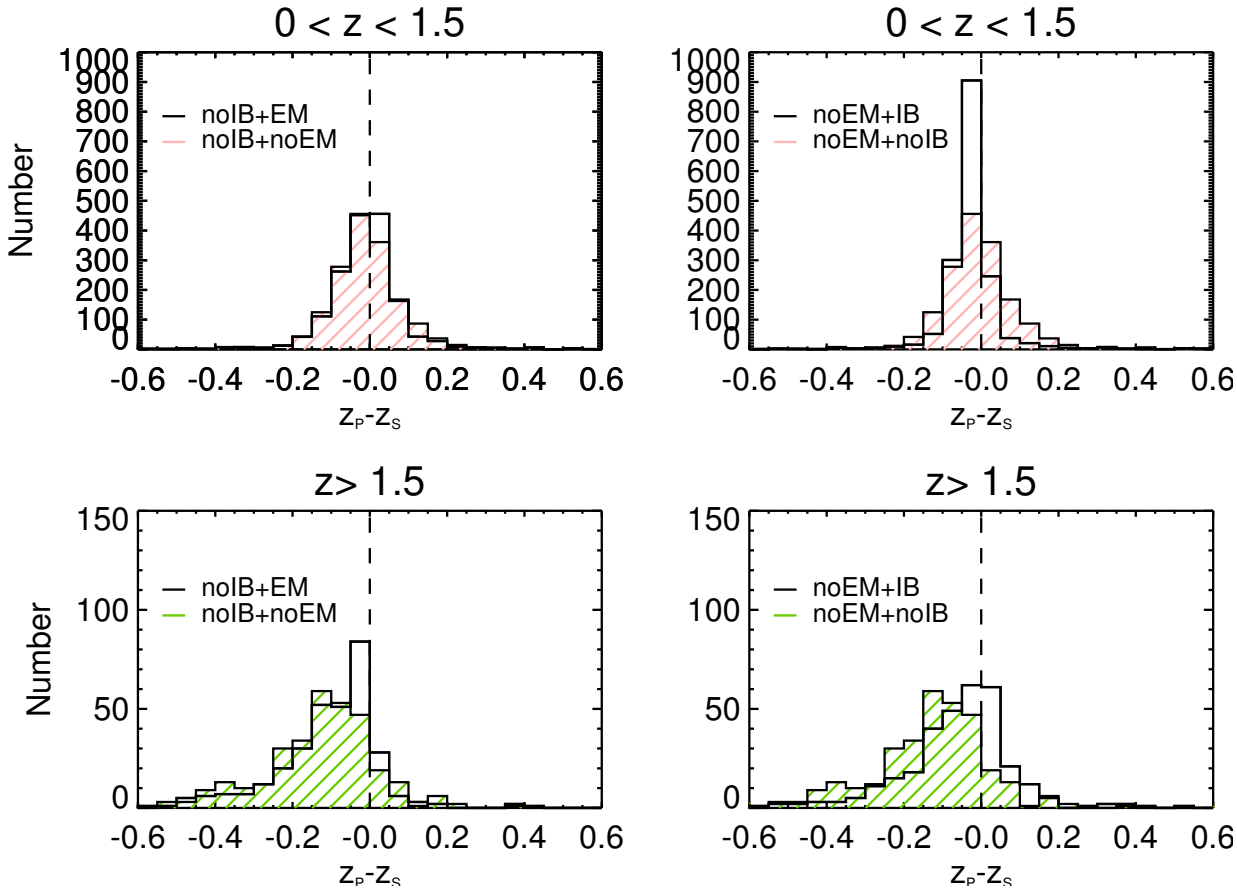


Figure 3.9: Comparisons of the $(z_{\rm p}-z_{\rm s})$ distribution with/without the use of IBs or emission lines (EMs) at $z < 1.5$ (upper panels) and $z > 1.5$ (lower panels).

In Figure 3.9, we compare the influence of using IBs and emission lines (EMs) on photo-z the overall results. The upper left plot shows that when IBs are not available, the use of emission lines in the templates produce only a marginal improvement on the accuracy at $z < 1.5$ (from 0.042 to 0.035). The lower left plot reveals that the lack of both IBs and EMs lead to systematically underestimated photo-z, specially at $z > 1.5$. Nevertheless, the upper right plot shows that when emission lines are not available, the photo-z result still can be greatly improved by using IBs at $z < 1.5$ (accuracy improves from 0.042 to 0.027). The lower right plot displays that using IBs helps to correct the underestimated photo-z result particularly at $z > 1.5$.

We quantified the comparisons in Table 3.3. Although the outlier fraction didn't change so much in four cases listed in the table, the photo-z accuracies are greatly improved when using IBs. This result reveals that using IB has bigger influence on photo-z quality than adding emission lines in the templates.

[]

Table 3.3: Influence of IB and EM on photo-z quality in Area 1

		$0 < z < 1.5$			$z > 1.5$		
IB	EM	bias_z	σ_{NMAD}	$\eta(\%)$	bias_z	σ_{NMAD}	$\eta(\%)$
yes	yes	0.001	0.011	3.62	-0.013	0.022	6.25
yes	no	-0.015	0.027	3.69	-0.017	0.032	5.81
no	yes	-0.007	0.035	4.92	-0.029	0.041	5.81
no	no	-0.005	0.042	4.68	-0.036	0.056	6.75

• Photometric Redshift Accuracy Beyond the Spectroscopic Limit

Using spec-z to estimate photo-z accuracy (as in Tables 3.1 and 3.6) is not representative of sources fainter than the spectroscopic limit. Quadri & Williams (2010) introduced a method for estimating photo-z accuracy based on the tendency of galaxies to cluster in space. Because of clustering, galaxies seen close to each other on the sky have a significant probability of being physically associated and having the same redshift. Therefore the distribution of photo-z differences[3] (Δz_p) of close pairs will show an excess at small redshift differences over the distribution for random pairs. This is seen in Figure 3.10.[4] The excess for close pairs with magnitude $J < 28$ fits a Gaussian with standard deviation $\sigma = 0.012$ in Area 1 and 0.010 in Areas 2+3. Because the width includes the scatter from both paired galaxies, the photo-z uncertainty for an individual object should be $\sigma/\sqrt{2}$ [5] . The pair test reveals that the faint sources without spec-z have photo-z accuracy similar to that of sources bright enough to have spectroscopic data. These values are given in Table 3.4 in each magnitude range.

Table 3.4: Photo-z Scatter from Pair Statistics

	Area 1	Areas 2+3
$J < 25$	0.008	0.007
$J < 26$	0.009	0.007
$J < 27$	0.009	0.007
$J < 28$	0.008	0.007

[3]Photo-z difference is defined as $\Delta z_p \equiv (z_{p,1} - z_{p,2})/(1 + z_{\text{mean}})$.

[4]Random pairs also show a noticeable peak at small Δz_p. This is not due to any systematic we can identify and may be due to the known large scale structure (Castellano et al., 2007; Salimbeni et al., 2009; Dehghan & Johnston-Hollitt, 2014) in the field.

[5] The close pair excess includes only objects with similar photo-z, so outliers are excluded in calculating σ here.

3.2 Photo-z for the X-ray selected sources

This work uses deep X-ray data to identify candidate AGNs. However, X-ray surveys as deep as the 4Ms-CDFS can also detect significant numbers of star-forming galaxies. The library must therefore include templates of normal galaxies, AGNs, and AGN/galaxy hybrids. In the following section, we illustrate the procedure we used to constructed our template libraries.

3.2.1 Template Library Methods

Luo et al. (2010) computed photo-z for sources in the 2Ms-CDFS by using the spectra of known sources as templates for SED fitting. Using the entire spectroscopic sample for template training gave an apparent accuracy ∼0.01 with almost no outliers. However, to get a more realistic estimation, Luo et al. (2010) randomly picked 75% of spectroscopic sample to train the SED templates and used the other 25% as the blind set. They repeated the process 20 times to perform statistically unbiased test, and obtained a more realistic accuracy of 0.059 and ∼9% outliers. This demonstrates how important the training sample is. For the same field, Cardamone et al. (2010b) combined normal galaxy templates with a type 1 QSO template on the fly when running SED fitting code for AGNs. The method gave accurate results ($\sigma_{\rm NMAD} \sim 0.01$) but a large outlier fraction ($\sim 12\%$).

Salvato et al. (2009, 2011, hereafter S09, S11) pursued a different approach for X-ray sources in the COSMOS field (Scoville et al., 2007) detected by *XMM* (Cappelluti et al., 2009) and by *Chandra* (Elvis et al., 2009). This involved: (1) correcting the photometry for variability when applicable; (2) separating the optical counterparts to the X-ray sources into two subgroups: point-like and/or variable sources in one and extended, constant sources in the other; (3) applying absolute magnitude priors to these two subgroups, assuming that the former are AGN-dominated while the latter are galaxy-dominated; (4) creating AGN-galaxy hybrids, using different libraries for the two subgroups rather than using only one type 1 AGN as in Cardamone et al. (2010b). This same procedure substantially reduced the fraction of outliers and gave higher accuracy than standard photo-z techniques when applied to X-ray sources in COSMOS. The procedure has also yielded reliable results for the Lockman Hole (Fotopoulou et al., 2012) and AEGIS fields (Nandra et al. 2015, submitted). S11 also verified the need for depth-dependent template libraries by showing that hybrids defined for *XMM*-COSMOS are not optimal for the deeper *Chandra*-COSMOS data. Even though the X-ray-faint *Chandra* sources are AGNs (i.e., $L_{\rm x} > 10^{42}$), since they are either host-dominated or obscured AGNs, normal galaxy templates with theoretical emission lines (Ilbert et al., 2009) gave better results for them than AGN-dominated templates.

3.2.2 Population-dependent SED Libraries

For this work, we constructed new hybrid templates following the procedure of S09 and S11 (Sec. 3.2.1). First we point out that the difference of the X-ray flux distributions between

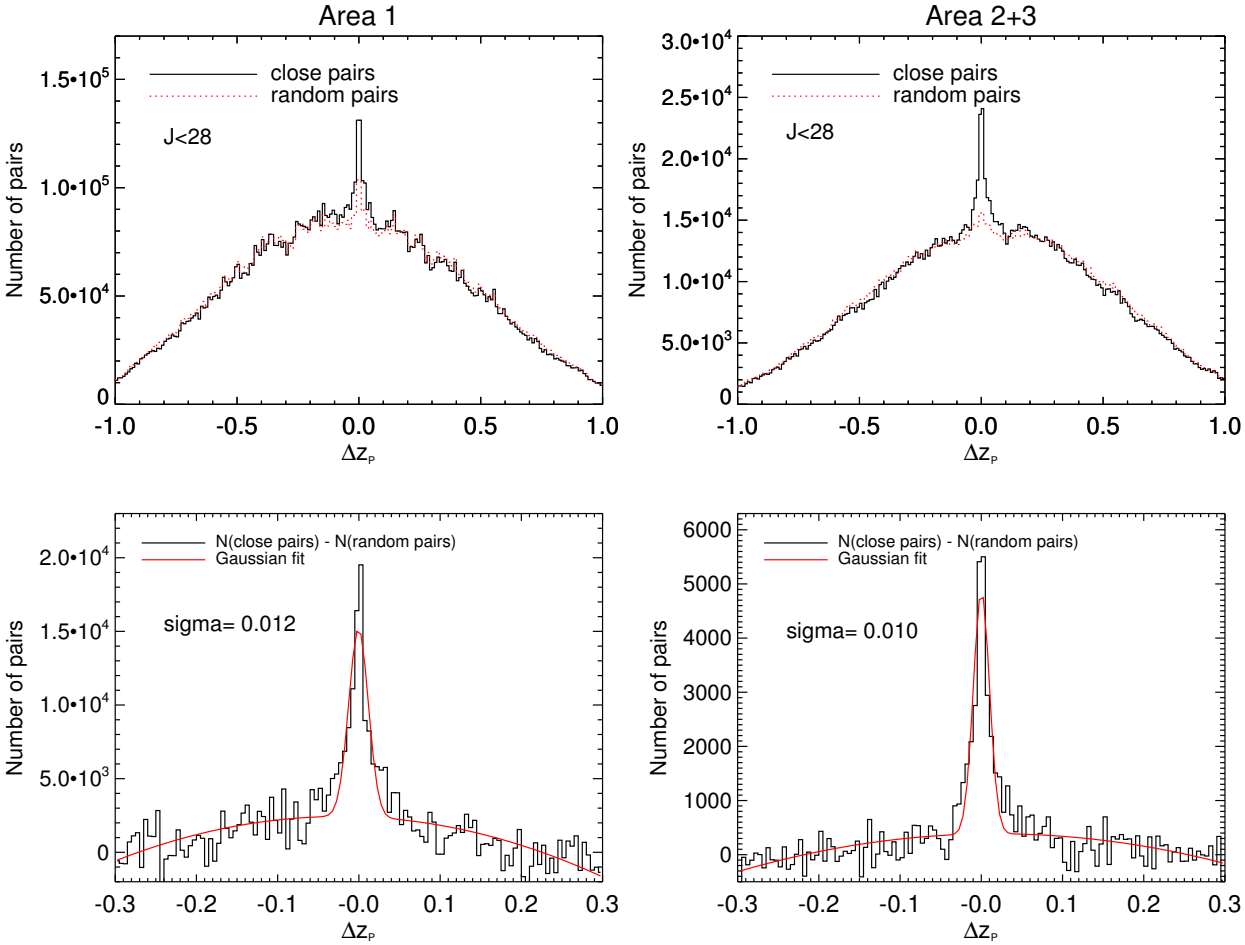


Figure 3.10: Distributions of Δz_p for pairs of galaxies. In the upper panels, red dotted lines represent differences for random pairs and black solid lines represent differences for pairs having angular separation $<15''$. Only galaxies with $J < 28$ are included. The lower panels show results for close pairs after subtracting the distributions for random pairs. Black lines show the observed Δz_p, and red solid lines show a Gaussian fit with standard deviation sigma as indicated in each panel. The left two panels are for Area 1 and the right two for Areas 2+3.

the two X-ray surveys used in this work (i.e., 4Ms-CDFS and 250ks-ECDFS) is even more extreme than that S11 (i.e., XMM-COSMOS and Chandra-COSMOS).

Fig. 3.11 shows the soft X-ray flux distributions of the 4Ms and 250ks sources, together with the distributions from the Chandra-COSMOS (Elvis et al., 2009) and XMM-COSMOS (Cappelluti et al., 2009). The left panel shows the distribution in numbers for each survey. Because of the sky coverage and the depth of the observations, most of the 4Ms-CDFS sources are located in the faint part of the flux distribution, which is opposite to the locus occupied by the shallower observations (e.g. XMM-/Chandra-COSMOS and 250ks-ECDFS). Normalizing by the total surveyed area[6] (see the right panel) reveals that the X-ray bright sources which are similar to the XMM-COSMOS sources are very rare in the 4Ms survey. This implies that the library of hybrids used in XMM-COSMOS is probably not representative of the 4Ms population. Based on thesea considerations, we need to build a new library for the fainter 4Ms-CDFS population.

Separate libraries were used for (a) point-like sources in Areas 1+2, (b) point-like sources in Area 3, and (c) extended sources in all Areas. Because the flux distribution of point-like sources in Area 3 is similar to that of the *XMM*-COSMOS field, library (b) for point-like sources in that area was the same as used by Salvato et al. (2009).

Note that we did not correct for variability as in S09 and S11. For the *XMM*-COSMOS field, Salvato et al. (2009) had multi-wavelength, multi-epoch observations spanning several years. About 1/4 of sources were seen to be variable. The lack of multi-epoch data for the CDFS/ECDFS means that we cannot detect variable objects and correct their photometry. However, these objects are a minor contributor to the X-ray population in the much smaller CDFS area (1/15 of XMM-COSMOS area where a larger fraction of type 1s contribute to the population). Therefore only a minor fraction of the Area 1 and 2 sources are likely to be variable. The major effect of being unable to correct for variability will be an increased outlier fraction rather than decreased photo-z accuracy (Salvato et al., 2009). Area 3, covered at 250 ks depth, is an intermediate case, and part of the photo-z inaccuracy there could be due to lack of variability correction. The spectroscopic testing in the respective Areas (Table 3.6) quantifies the outlier fractions and the inaccuracies resulting from all causes.

3.2.3 Constructing AGN-Galaxy Hybrid Templates

AGN-Galaxy hybrids were constructed by combining an AGN template with normal galaxy templates in various proportions. The AGN templates were QSO2 and modified TQSO1 (i.e., pl-TQSO1 built in Salvato et al. 2009 by adding power-law UV component to TQSO1) originally from Polletta et al. (2007). The galaxy templates are derived from Bender et al. (2001) with different intensities of empirical emission lines corresponding to the levels of star formation (see Fig. 3.12). Including these empirical emission lines in our hybrids is the major improvement on the SEDs built by Salvato et al. (2009), which used templates of galaxies with few or no emission line contributions.

[6]This is not the Log N-Log S

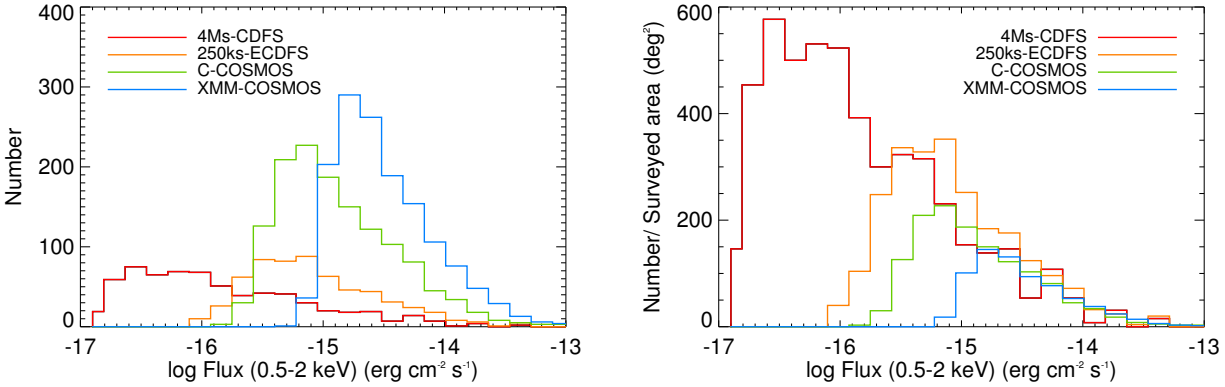


Figure 3.11: Soft X-ray flux distributions in numbers (left) and source densities (right). Histograms show distributions for the 4Ms-CDFS (Areas 1 and 2), 250ks-ECDFS (Area 3), and comparison surveys *Chandra*-COSMOS (Elvis et al., 2009) and *XMM*-COSMOS (Cappelluti et al., 2009) surveys as indicated in the legend.

The 32 galaxy templates were constructed by first sorting galaxies of known spec-z in the FORS Deep Field (Appenzeller et al., 2004; Gabasch et al., 2004) iteratively into 32 bins of similar spectral shape. Broad-band fluxes from the U-band to the K-band of typically ten galaxies at different redshifts were combined to obtain one broadband template covering as wide a wavelength range as possible. These broad-band empirical templates were fitted by a combination of model spectral energy distributions from Bruzual & Charlot (2003) and Maraston (2005) and empirical spectra from Noll et al. (2004) to obtain "semi-empirical" templates with spectral resolution $R \sim 1000$. The method covered wavelengths from ∼60 nm to 2.5 μm. Figure 3.13 illustrates two of the templates in use. These semi-empirical galaxy SEDs have already been successfully used by Gabasch et al. (2004), Drory et al. (2005), and Feulner et al. (2005).

In order to combine the AGN templates with galaxy templates, we normalized both AGN and galaxy templates at 5500 Å, then combined them with the AGN-to-galaxy ratio changing from 1:9 to 9:1. (See examples in Fig. 3.14.) In total, 576 hybrids were created this way.

Even if these 576 hybrids are representative, they can not be used all together because of the degeneracy issue. Thus, the library should be representative but also limited. In order to reduce the number of templates, we looked at the optical images of X-ray selected AGNs, and split the sources into extended and point-like subgroups. We visually inspected WFC3/H-band images for Area 1 and the MUSYC/BVR images for Areas 2 and 3. The extended sources were assumed to be host-dominated, and being seen as extended means they are likely to be at low redshift. For these sources, we applied an absolute magnitude prior $-24 < M_B < -8$ and used templates with at most a small AGN fraction. Point-like sources are usually AGN-dominated and can be at any redshift. We therefore applied a

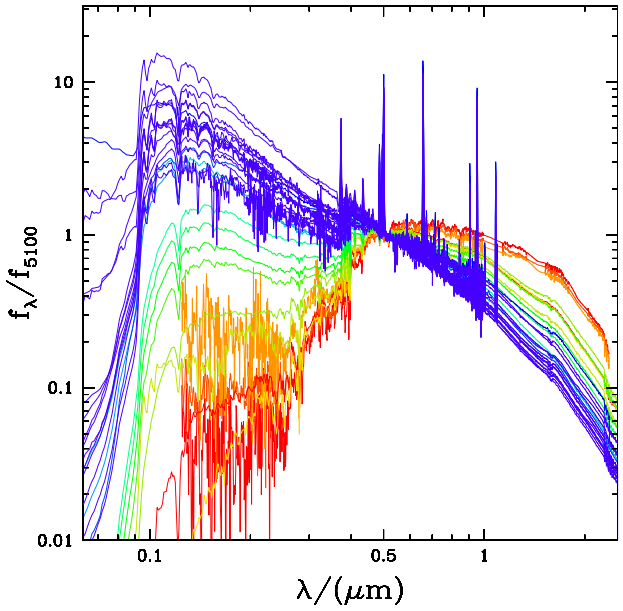


Figure 3.12: Galaxy templates from Bender et al. (2001), color coded as a function of activity from the redder passive galaxies to the bluer strongly star forming objects.

prior $-30 < M_B < -20$ to these and used hybrid AGN-galaxy templates.

To train the hybrids, we then randomly chose $\sim 25\%$ of the reliable spectroscopic sample from the 4Ms-CDFS X-ray catalogs, (52 extended sources and 62 point-like sources) to train the hybrids. The remaining objects were used for unbiased testing of the results.

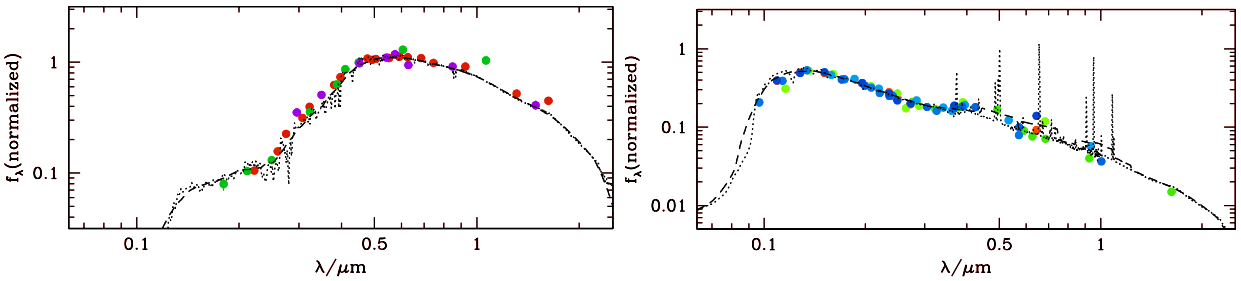
As Figure 2.4 has shown, the training samples are well distributed over the entire ranges of redshift and magnitude. We treated the extended and point-like sources separately, fixing the redshift at the spectroscopically defined value and choosing the templates most frequently selected to represent the training sources. After several iterations, we obtained the libraries used for the extended sources (Lib-EXT: 31 hybrids + 17 galaxy templates, see Fig. 3.15) and for the point-like sources (Lib-PT: 30 hybrids, see Fig. 3.16).

Table 3.5 lists the templates in Lib-EXT and Lib-PT. Names with "-TQSO1-" or "-QSO2-" indicate the AGN component used. The number following indicates the fractional AGN contribution in the hybrid. For example, the template "s050-8-TQSO1-2" contains 80% galaxy (s050) and 20% of AGN (TQSO1). The templates without TQSO1 and QSO2 are pure galaxies with different levels of star formation (see Bender et al. 2001 for details).

3.2.4 Resuts

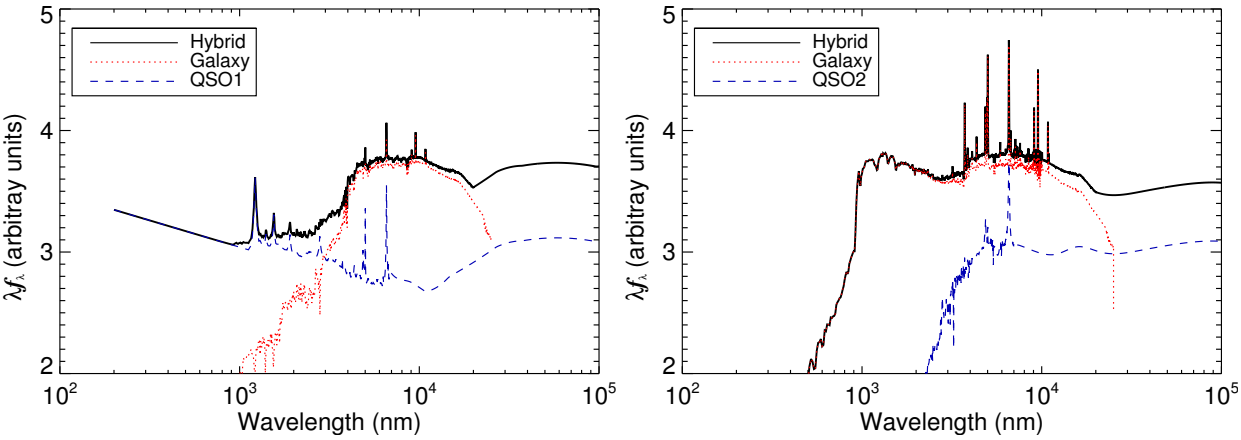
• Area 1 and 2

In Area 1, where $\text{TFIT}_{\text{CANDELS+IB}}$ photometry and high-resolution space-based images are available, the photo-z for X-ray sources (Table 3.6) are as accurate as those for normal galaxies (Table 3.1). Remarkably, the outlier fraction is actually lower for the X-ray sources than for normal galaxies. Excellent photo-z quality is maintained even for $z > 1.5$. Fig-



(a) (b)

Figure 3.13: Galaxy templates in use for photometric redshift estimation. Colored symbols represent broad-band flux densities of individual galaxies with known spec-z. The short-dashed line shows the $R \sim 1000$ spectral template best fitting the broadband flux densities after smoothing to the broad-band resolution. The smoothed template is shown as a long-dashed line.



(a) (b)

Figure 3.14: Two examples of hybrid templates. Red lines show the galaxy contribution, blue lines the AGN contribution, and black lines the sum. The left panel shows a hybrid comprised of 10% Type 1 AGN and 90% galaxy. The right panel shows a hybrid with 30% Type 2 QSO and 70% starburst galaxy.

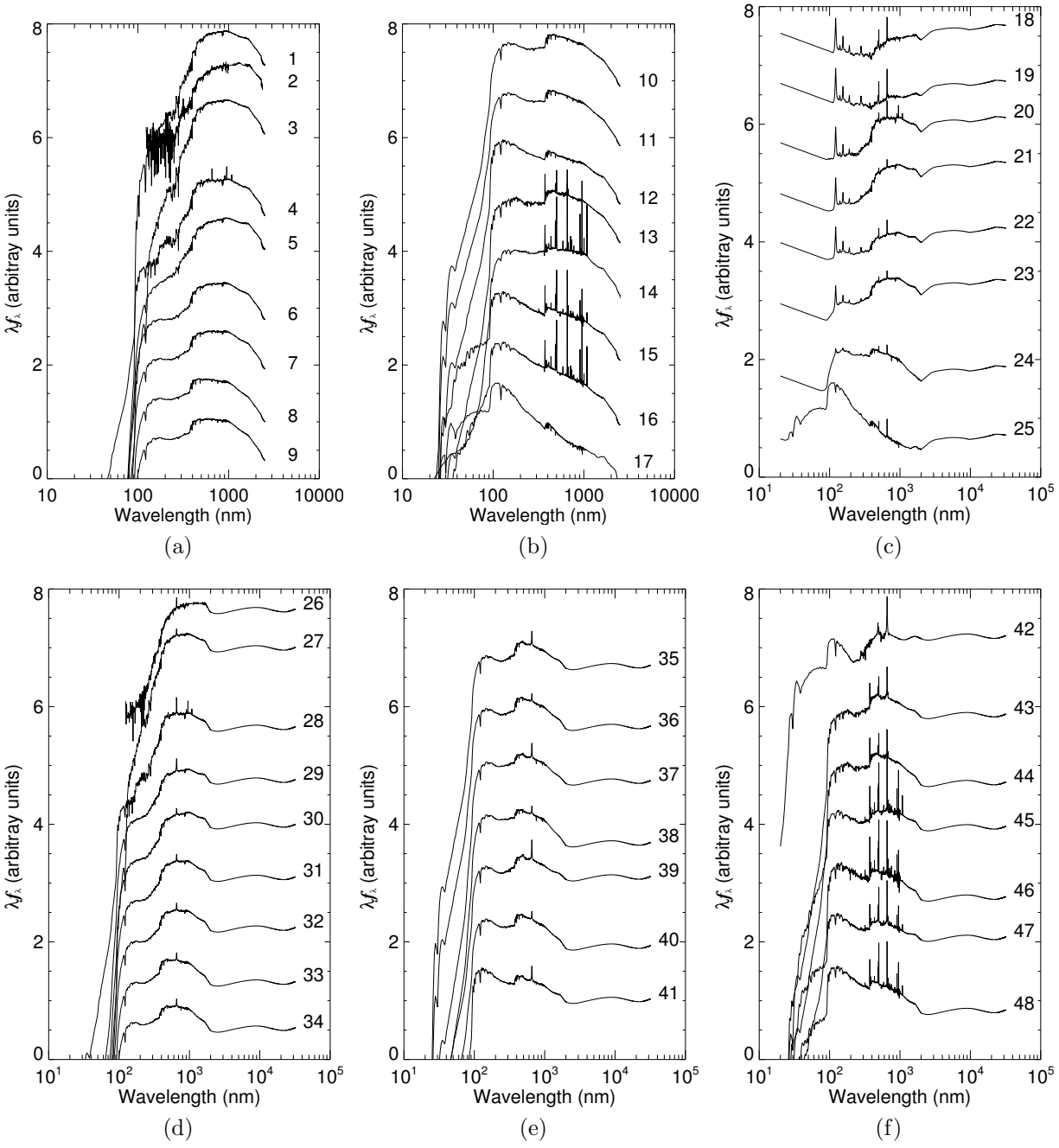
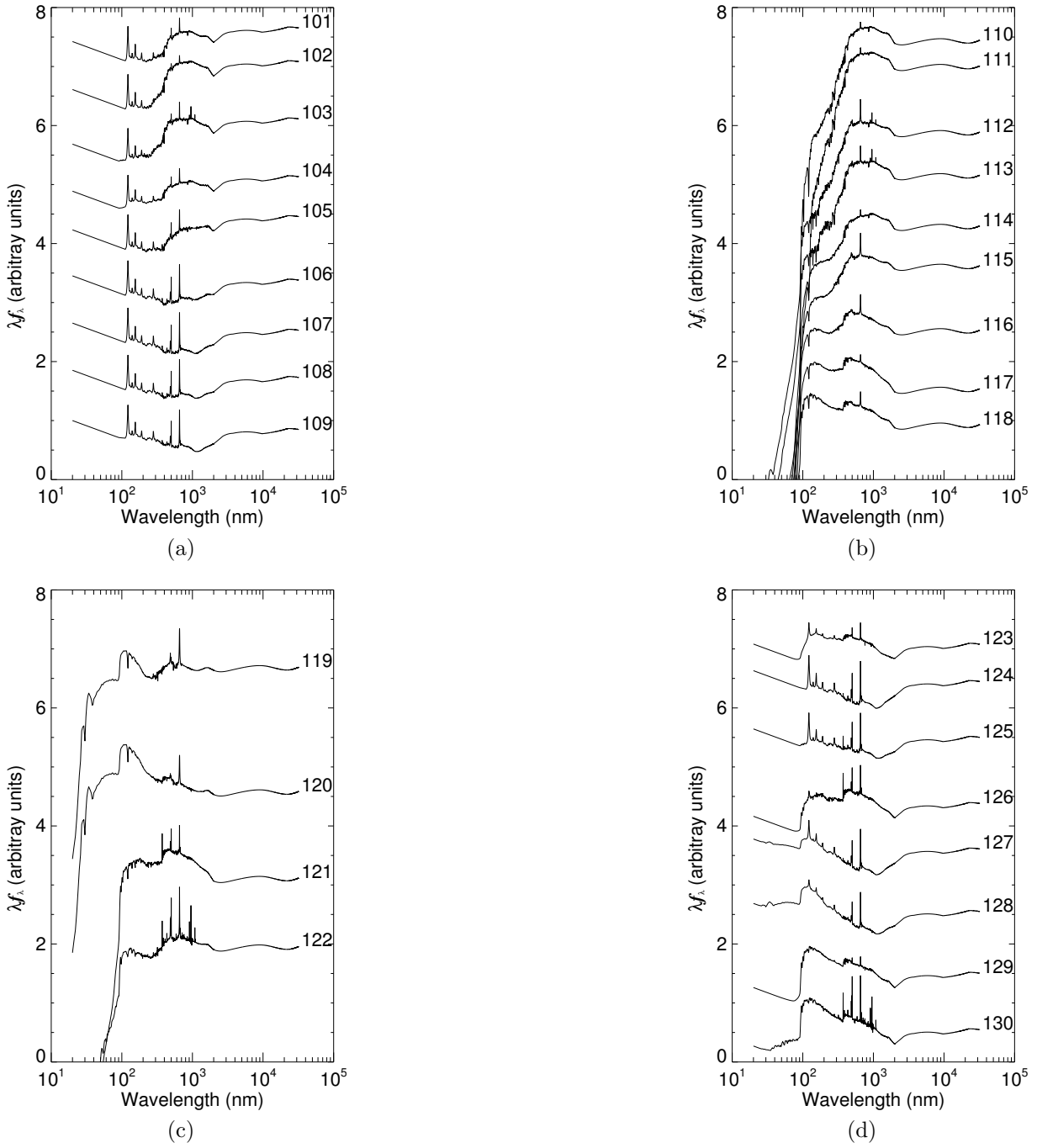


Figure 3.15: SEDs for all templates in Lib-EXT.



(a) (b) (c) (d)

Figure 3.16: SEDs for all templaes in Lib-PT.

Table 3.5: The list of AGN-galaxy hybrids in Lib-EXT and Lib-PT.

Lib-EXT		Lib-PT	
No.	Template	No.	Template
1	mod-e	101	e-8-TQSO1-2
2	manucci-sbc	102	s010-9-TQSO1-1
3	mod-s010	103	s020-9-TQSO1-1
4	mod-s020	104	s050-8-TQSO1-2
5	mod-s030	105	sac-7-TQSO1-3
6	mod-s070	106	ec-3-TQSO1-7
7	mod-s090	107	sac-2-TQSO1-8
8	mod-s120	108	s010-3-TQSO1-7
9	mod-s150	109	s180-3-TQSO1-7
10	mod-s200	110	e-9-QSO2-1
11	mod-s400	111	s010-9-QSO2-1
12	mod-s500	112	s020-7-QSO2-3
13	mod-fdf4	113	s020-9-QSO2-1
14	mod-s210	114	s050-9-QSO2-1
15	mod-s670	115	s090-6-QSO2-4
16	mod-s700	116	s200-7-QSO2-3
17	mod-s800	117	s400-9-QSO2-1
18	ec-6-TQSO1-4	118	s500-8-QSO2-2
19	sac-5-TQSO1-5	119	s800-2-QSO2-8
20	s020-9-TQSO1-1	120	s800-5-QSO2-5
21	s030-9-TQSO1-1	121	fdf4-9-QSO2-1
22	s050-8-TQSO1-2	122	s230-5-QSO2-5
23	s070-9-TQSO1-1	123	s250-8-TQSO1-2
24	s250-9-TQSO1-1	124	s250-1-TQSO1-9
25	s800-8-TQSO1-2	125	fdf4-4-TQSO1-6
26	sac-9-QSO2-1	126	fdf4-9-TQSO1-1
27	s010-9-QSO2-1	127	s800-2-TQSO1-8
28	s020-9-QSO2-1	128	s800-4-TQSO1-6
29	s050-8-QSO2-2	129	s500-9-TQSO1-1
30	s050-9-QSO2-1	130	s670-9-TQSO1-1
31	s070-9-QSO2-1		
32	s090-9-QSO2-1		
33	s120-9-QSO2-1		
34	s180-9-QSO2-1		
35	s200-8-QSO2-2		
36	s200-9-QSO2-1		
37	s250-8-QSO2-2		
38	s250-9-QSO2-1		
39	s400-7-QSO2-3		
40	s400-9-QSO2-1		
41	s500-8-QSO2-2		
42	s800-1-QSO2-9		
43	fdf4-7-QSO2-3		
44	fdf4-9-QSO2-1		
45	s230-8-QSO2-2		
46	s650-9-QSO2-1		
47	s670-6-QSO2-4		
48	s670-9-QSO2-1		

ure 3.17 shows that the results are largely attributable to the deblended IBs data and the WFC3 data with their depth and high resolution. The WFC3 NIR data are much deeper and have smaller flux errors than the MUSYC ground-based data (as shown in Figure 3.5). With this superior data set, we are able not only to obtain more accurate photo-z, but also to reduce the degeneracy while running the fitting with such small errors. Instead of using ground-based data (e.g., MUSYC+TENIS), the use of $\mathrm{TFIT}_{\mathrm{CANDELS+IB}}$ catalog allows us to reduce the outlier fraction by a factor of 3. The improvement is especially great for the $R > 23$ and $z > 1.5$ sources. The outlier fractions decrease from 6.3% to 2.1% for faint sources and from 12.8% to 2.6% for high-redshift sources. Comparison with Area 2 also confirms the importance of the WFC3 data. Without these data, photo-z accuracy deteriorates only slightly (Table 3.6), but the outlier fraction triples. Most of the outlier increase comes from the $R > 23$ and $z > 1.5$ subsets. (There are only two sources with $H > 23$, and numerical results for that bin are meaningless.)

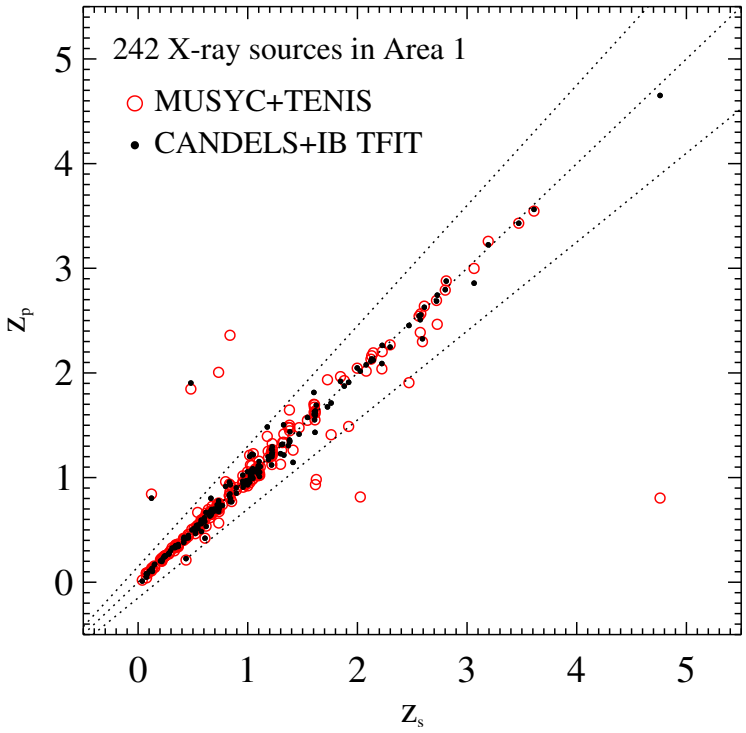


Figure 3.17: Comparison of photo-z to spec-z with and without $\mathrm{TFIT}_{\mathrm{CANDELS+IB}}$ photometry. Filled points show results for 242 X-ray sources from the Cardamone et al. (2010b) catalog using the full $\mathrm{TFIT}_{\mathrm{CANDELS+IB}}$ dataset. Open circles show results for the same sources using only the MUSYC+TENIS data.

• Area 3

Area 3 has a larger fraction of outliers than either of the other two Areas, though accuracy for the non-outliers is little worse than in Areas 1 and 2 (Table 3.6). Three effects probably contribute to the larger fraction of outliers. One is shallower photometry (i.e., large flux uncertainty for faint objects) at the border of the field (Fig. 2.1), leading

to larger errors. Second, the X-ray coverage is shallower in the larger Area 3, thus the fraction of varying Type 1 AGN is presumably higher. The lack of variability correction will therefore have a larger effect. This is likely exacerbated by the third effect, having to use ground-based images rather than higher-resolution images for classifying sources as point-like or extended. In Area 1, about 30% of sources are classified point-like using WFC3 but extended on a ground-based image, **due to the large uncertainty on the morphology classification according to the point spread function (psf) of low resolution of the images.** Using the template library for the extended sources rather than for point-like classification would have doubled the outlier fraction.

• **Global result**

Overall, in Area 1+2+3, we obtained a photo-z accuracy of 0.013 with the fraction of outliers = 5.3% (Figure 3.18) for 566 X-ray-detected sources which have reliable spec-z. See the detailed photo-z quality in each redshift and magnitude bin in Table 3.6.

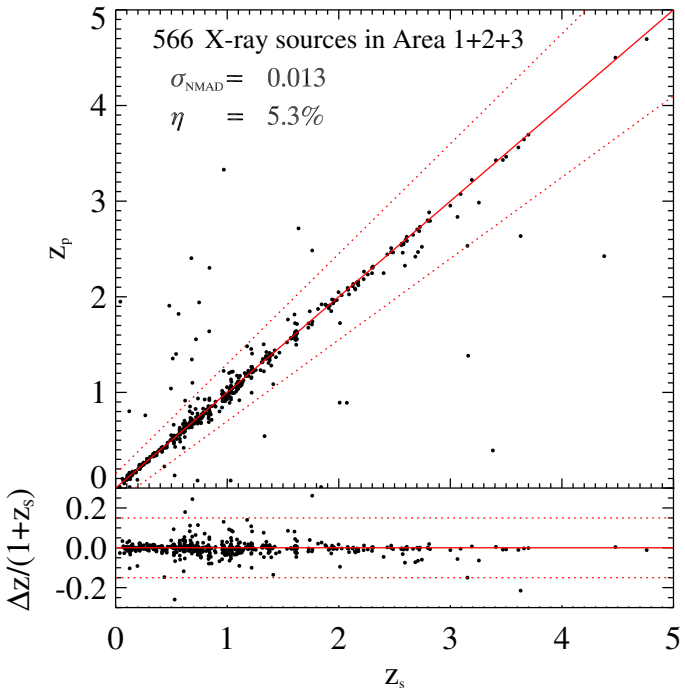


Figure 3.18: Comparison of photo-z to spec-z for all X-ray sources in Areas 1+2+3.

Furthermore, in order to identify possible outliers among the sources without spec-z, we look at the distribution of observed-frame X-ray luminosity as a function of redshift. In Figure 3.19, all of the X-ray sources either from 4Ms or 250ks surveys are located above the lower limit of luminosity. However, three sources with apparent extreme redshift are probably outliers. They are located on the edge of the optical images and have unreliable or non-existent MUSYC photometry, leaving only six photometric data points (from the TENIS catalog). In this situation, the limited photometry and large photometric errors

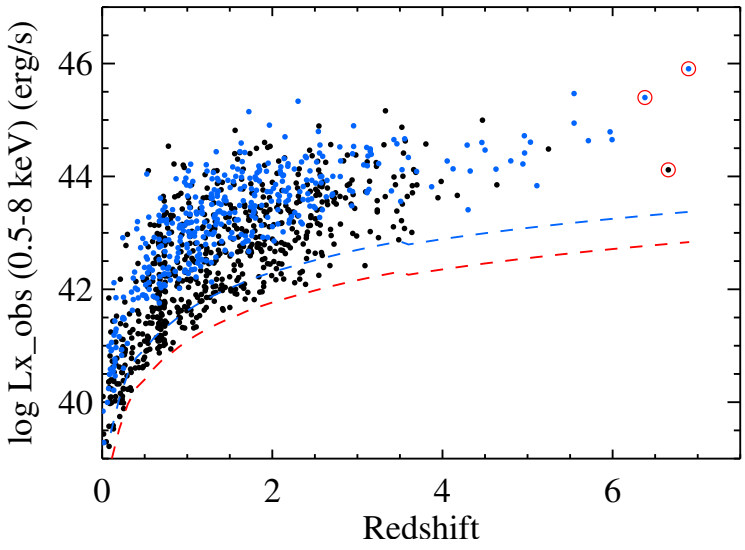


Figure 3.19: Distribution of 0.5–8 keV observed-frame X-ray luminosity as a function of redshift for all X-ray sources. Redshifts are spec-z if available and otherwise photo-z. Black dots are the X-ray sources from the 4Ms CDFS survey with the flux limit 3.2×10^{-17} erg cm^{-2} s^{-1} (red dashed line). Blue dots are the X-ray sources from the 250ks ECDFS survey with the flux limit of $\sim 10^{-16}$ erg cm^{-2} s^{-1} (blue dashed line). Red open circles indicate the three high-redshift candidates whose photo-z should take account of the probability distribution function rather than using only pure the peak value of p(z).

can not provide sufficient constraints. Particularly in these cases, the photo-z should not be taken as a pure number, but instead the redshift probability distribution function should be considered.

Table 3.6: Photo-z Quality for X-ray Sources

	Area 1				Area 2				Area 3				Area 1+2+3			
	N	$bias_z$	$\sigma_{\rm NMAD}$	$\eta(\%)$	N	$\rm bias_z$	$\sigma_{\rm NMAD}$	$\eta(\%)$	N	$\rm bias_z$	$\sigma_{\rm NMAD}$	$\eta(\%)$	N	$\rm bias_z$	$\sigma_{\rm NMAD}$	$\eta(\%)$
Total	300	-0.002	0.012	2.00	104	-0.003	0.014	5.77	162	-0.003	0.015	11.11	566	-0.002	0.013	5.30
$R < 23$	172	-0.003	0.010	1.16	80	-0.000	0.014	5.00	115	0.002	0.011	8.70	367	-0.001	0.011	4.36
$R > 23$	128	0.001	0.020	3.12	24	-0.014	0.016	8.33	47	-0.017	0.026	17.02	199	-0.004	0.021	7.04
$H < 23$	279	-0.002	0.012	1.79	102	-0.003	0.014	5.88	72	-0.006	0.015	6.94	453	-0.003	0.013	3.53
$H > 23$	21	0.011	0.011	4.76	2	-0.010	0.027	0.00	90	-0.001	0.016	14.44	113	0.001	0.014	12.39
$z < 1.5$	240	-0.001	0.012	2.08	86	-0.001	0.015	4.65	124	-0.002	0.018	9.68	450	-0.001	0.013	4.67
$z > 1.5$	60	-0.003	0.014	1.67	18	-0.018	0.012	11.11	38	-0.010	0.009	15.79	116	-0.007	0.012	7.76
Photo-z quality for the training set of AGN/galaxy hybrids													78	0.006	0.015	1.3

3.2.5 Discussion

• Testing Libraries for the X-ray Population

Because of the different X-ray populations in the 4Ms-CDFS and 250ks-ECDFS surveys, we adopted different libraries for point-like sources in Areas 1+2 and Area 3. For the sake of template comparison, we tried using the Area 3 library ("S09", i.e., templates tuned for bright AGNs) to calculate photo-z for point-like sources in Areas 1+2. The fraction of outliers increased from 5.3% to 15%, and the accuracy became two times worse than achieved with the preferred library. Even for $R < 23$ sources, $\sigma_{\rm NMAD}$ went from 0.011 with the proper templates to 0.016 with the old ones. For $R > 23$ galaxies, the deterioration was from 0.027 to 0.059. In Area 3, on the other hand, using the new templates instead of the S09 ones made photo-z slightly worse: for $R < 23$, $\sigma_{\rm NMAD}$ was 0.009 for the new and 0.008 for the S09 libraries. For $R > 23$, accuracies were 0.025 and 0.017 respectively. Moreover, the $\mathrm{bias}_z = -0.014$ using the S09 library but increased to $\mathrm{bias}_z = -0.031$ with the new library. The better performance of the S09 library in Area 3 can be understood because the population of point-like X-ray sources in the 250k-ECDFS is similar to the *XMM*-COSMOS population, and the S09 library is more suitable for counterparts of such bright X-ray sources.

• Impact of UV data

UV emission from accretion disks around supermassive black holes makes type 1 AGNs distinguishable from normal galaxies. Figure 3.20 illustrates that the slope of the continuum in the UV waveband changes with different types of AGNs. For the type 2 AGN (Sey2 and QSO2), as the UV light is obscured by the dust, the SED is similar to the normal galaxy. However for the type 1 AGN, the strong UV excess appears and have a distinct power-low feature. Therefore, to improve photo-z computed via SED fitting for type 1 AGNs, UV data are crucial, because it helps in breaking the degeneracy when only optical data is available. The presence of UV data imposes a lower redshift, avoiding putting the sources among the high-z population. This point is well illustrated in Figure 3.21, in that most of the sources with overestimated photo-z are corrected after adding UV data.

To demonstrate this further we compared photo-z for AGNs obtained with and without photometry in the UV bands. About 25% of all the X-ray detected sources in Areas 1+2+3 have UV data available from GALEX. Among these, 221 sources have spectroscopy available and were used as our test sample. As expected, for the optically extended sources, where the host dominates the emission, there is very little difference in accuracy and fraction of outliers whether UV data are included or not. For the 51 point-like (i.e., AGN-dominated) spectroscopic sources, adding the UV data halves the number of outliers (from 23.5% to 11.8%) though with only modest improvement in the already good accuracy (from 0.013 to 0.011). Among the five remaining outliers (see Fig. 3.21), two are faint (mag > 23) in the UV, and three are close to other sources with the UV flux blended in the 10″ GALEX aperture. Deblending the GALEX photometry with TFIT as in the other bands could perhaps improve these cases. For 170 extended sources (i.e., host-dominated),

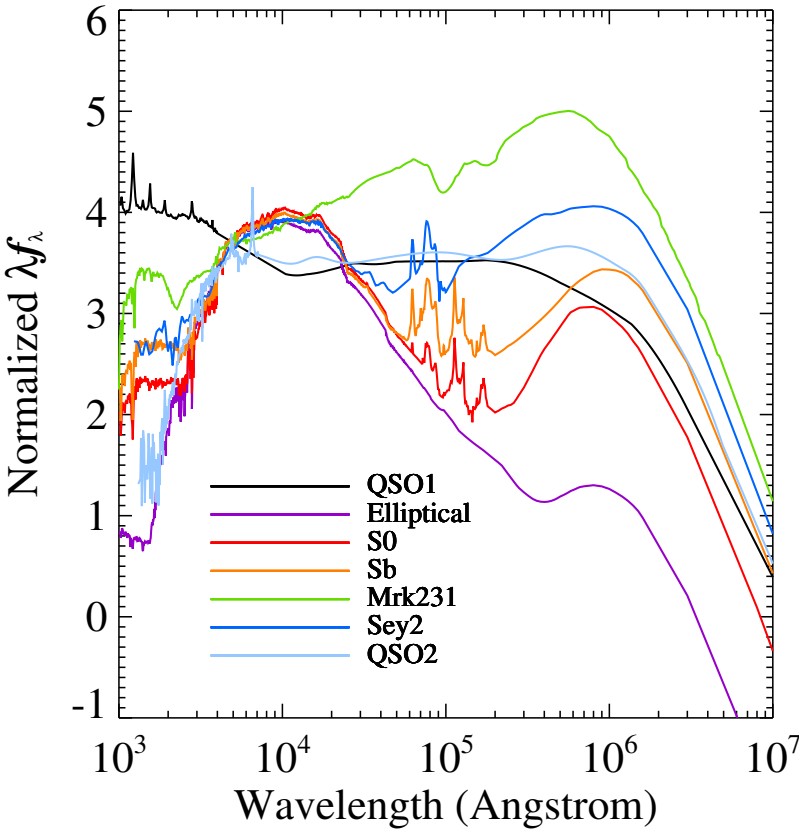


Figure 3.20: Spectral energy distributions for typical galaxies and AGNs with the flux normalized at 5000 Å. Template spectra are taken from Polletta et al. (2007).

as we expected, including UV data marginally decreases $\sigma_{\rm NMAD}$ from 0.013 to 0.012 and η from 5.9% to 5.3%.

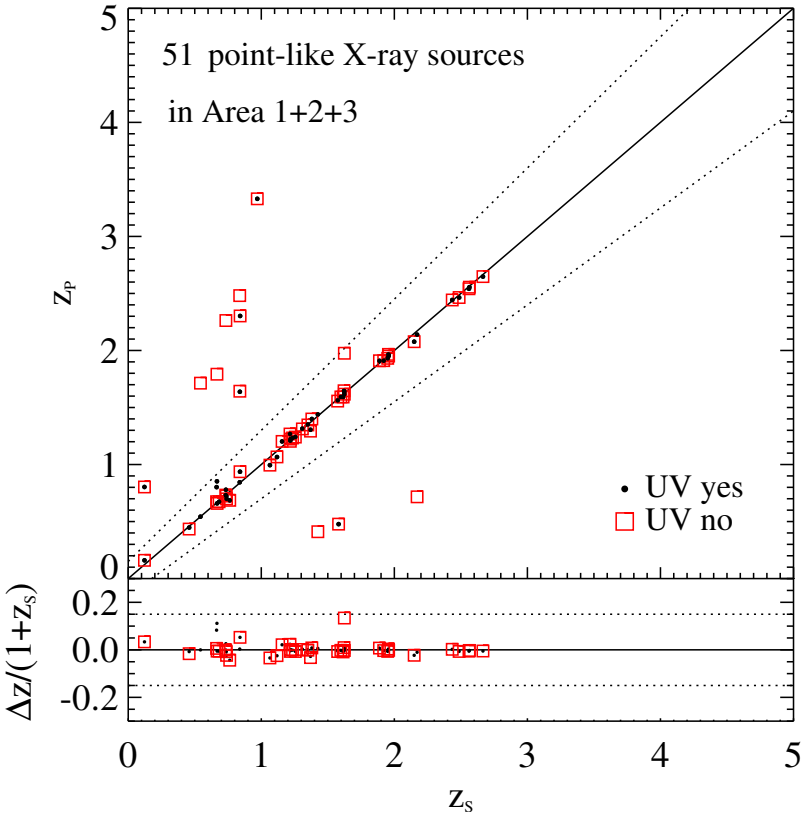


Figure 3.21: Comparison between photo-z computed with and without UV data for the X-ray sources that are point-like in H-band. All 51 available sources in Areas 1+2+3 are plotted. Black dots indicate photo-z computed with UV data, and red squares indicate photo-z computed without UV data.

Chapter 4

Color-magnitude relations for AGNs and galaxies

The relation between rest-frame optical colors ($U-V/U-B$) and absolute magnitudes (V-/B-band) is a useful probe to trace the stellar populations in galaxies (Bell et al., 2004), because optical colors are affected by dust and star formation in galaxies. Thus using the color-magnitude diagram (CMD) to study the galaxy and AGN host colors distributions can provides an important clue to AGN and galaxy (co)evolution.

In normal galaxies, the $(U-V)$ colors show a strong bimodality which divides galaxies into passive (i.e., red sequence) and star-forming (i.e., blue cloud) galaxies (Bell et al., 2004; Wuyts et al., 2007; Williams et al., 2009). On the contrary, for X-ray-selected AGNs, many works found that no clear color bimodality is presented in CMD (e.g., Nandra et al., 2007; Silverman et al., 2008a; Cardamone et al., 2010a; Xue et al., 2010). They revealed that AGN host galaxy have intermediate color between red and blue at $z \leq 1$, called the green valley, which is interpreted as the evolutionary transition from star-forming to passive galaxy. For instance, Cardamone et al. (2010a) found that few AGN host galaxies are located in the intermediate area of the green valley. They considered this transition as the observational evidence of AGN feedback which heat up galaxy gas and shutdown the star-forming activity in the host galaxy.

However, it is always a huge challenge to get a precise AGN host galaxy color because: (1) both early-type galaxies and dusty galaxies can present red colors, and (2) both AGN and host can contribute to the optical emission in different proportion. To solve this two problems, we make corrections for dust extinction and AGN contribution based on the well-trained AGN-galaxy hybrids for our SED-fitting, de-reddening the host galaxy color and subtracting the flux from the central AGNs.

Furthermore, due to the poor quality of photo-z in the early days, the CMD study was usually restricted to spectroscopic sample (e.g., Nandra et al., 2007; Silverman et al., 2008a). In recent years, photo-z quality has been improved greatly for galaxies and AGNs (e.g., Ilbert et al., 2009; Salvato et al., 2009, 2011) as the intermediate-band photometry becoming available, such as the COMBO-17 (Wolf et al., 2003) and the MUSYC survey (Cardamone et al., 2010b). In our work, in addition to the intermediate-band photometry,

we utilize the deep space-based photometry to obtain highly accurate photo-z (Hsu et al., 2014). Thus we can push the study of CMD to the more distant Universe compared to previous works.

In this chapter, we will describe the procedures of de-reddening and AGN/galaxy decomposition in detail, and discuss the CMD for both AGNs and galaxies in the CANDELS region (i.e., Area 1)

4.1 Dust correction and AGN decomposition

4.1.1 Dust de-reddening

Red optical colors can be caused by dusty star formation galaxies or an old stellar population in passive galaxies. Many of the green-valley sources in the color-magnitude diagram are found to be dust-reddened blue-cloud objects (Cowie & Barger, 2008; Brammer et al., 2009), where the color $(U-V)$ is overestimated due to dust. Cardamone et al. (2010a) also found that roughly half of the red-sequence AGNs have bluer colors after de-reddening.

We correct for dust extinction for the rest-frame colors $(U-V)$ based on the A_λ (shown in Equation 4.1) derived from the SED-fitting results, assuming the extinction law of Calzetti et al. (2000) or Prevot et al. (1984) .

$$f(\lambda)_{\rm ext} = f(\lambda)_{\rm int} \times 10^{-0.4A_\lambda} \tag{4.1}$$

where$f(\lambda)_{\rm int}$ is the intrinsic flux, and $f(\lambda)_{\rm ext}$ is the observed flux with dust extinction. A_λ is the extinction at wavelength λ, related to reddening excess $E(B-V)$ and the extinction curve $k(\lambda)$. We obtained the value of A_λ for each source by the output :

$$A_\lambda = k(\lambda) \times E(B-V) \tag{4.2}$$

Figure 4.1 shows four reddening curves, $k(\lambda)$, as a function of redshift used in our work (i.e., Prevot et al. 1984 and Calzetti et al. 2000 plus two more modified curves with bumps). We use these four extinction laws in the SED fitting for galaxies and AGNs. The fitting result of each source provides us the best-fit extinction law and the value of $E(B-V)$. Therefore we can use these information to compute A_λ at the effective wavelength of U-band and V-band (i.e., 3778Å and 5737Å, respectively).

4.1.2 AGN-host decomposition

At low redshift ($z < 1$), the AGN/host decomposition is often performed by directly subtracting AGN flux contribution from images using tools like $GALFIT$ (Peng et al., 2002). However at high redshift, It is not possible to separate the AGN and host contribution because the image resolution is not sufficient to distinguish between them. In our work, we utilize the output information from the SED-fitting to do the AGN/host decomposition for the X-ray sources in CANDELS region. The parameters we used in the following

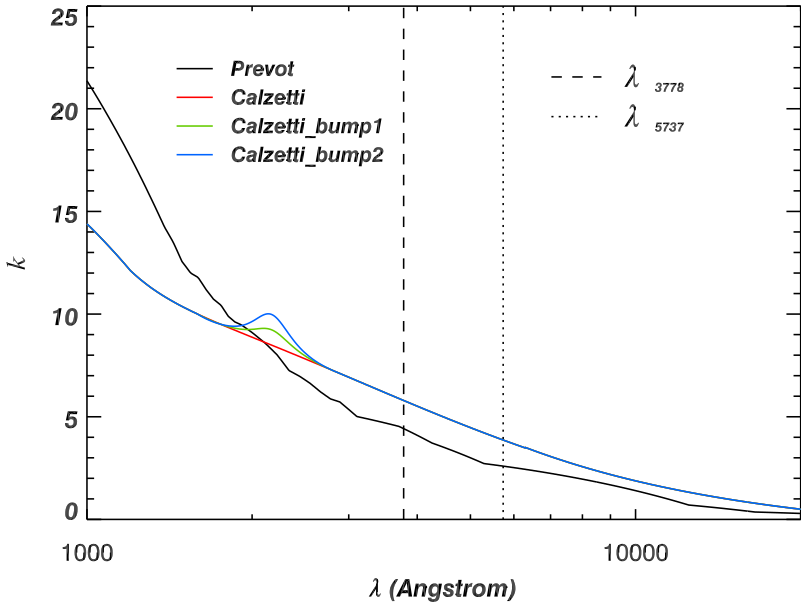


Figure 4.1: Four extinction laws used in our SED fitting.

procedures are: (a) the best-fit template, (b) the best-fit extinction law, (c) E(B-V), and (d) scaling factor. With these parameters, we can disentangle the contribution from AGN, host, and dust, at any λ to the SED of the sources. The procedure that we followed is described below:

1. First, according to the parameter (a), the best-fit hybrid/template for each source shows what fraction of AGN and galaxy (denoted by A and B, respectively) contribute to the hybrid (see Table 3.5). For example, if the name of the best-fit hybrid is "s800-1-QSO2-9", we know that the hybrid consists of 90% (i.e., $A = 0.9$) of QSO2 and 10% (i.e., $B = 0.1$) of star-forming galaxy (s800) normalizing at rest-frame 5500Å.

2. Second, from parameters (b) and (c), we can infer the best-fit extinction law and E(B-V) that needs to be used to de-redden the host and the AGN fluxes separately. We normalized the observed-frame AGN and galaxy template at 5500Å $\times$ (1 + z), and multiplied by the fraction of A and B separately as below.

$$f(\lambda)^{\mathrm{AGN,ext}}_{\mathrm{temp}} = A \cdot f(\lambda)^{\mathrm{AGN,ext}}_{\mathrm{temp,norm}} \tag{4.3a}$$
$$f(\lambda)^{\mathrm{host,ext}}_{\mathrm{temp}} = B \cdot f(\lambda)^{\mathrm{host,ext}}_{\mathrm{temp,norm}} \tag{4.3b}$$

By combining the AGN and galaxy fluxes, we can have total flux $f(\lambda)^{\mathrm{total,ext}}_{\mathrm{temp}} = f(\lambda)^{\mathrm{AGN,ext}}_{\mathrm{temp}} + f(\lambda)^{\mathrm{host,ext}}_{\mathrm{temp}}$. Written together with the dust extinction from Equation

4.1:

$$f(\lambda)^{\rm total,ext}_{\rm temp} = f(\lambda)^{\rm AGN,int}_{\rm temp} \times 10^{-0.4A_\lambda} + f(\lambda)^{\rm host,int}_{\rm temp} \times 10^{-0.4A_\lambda} \tag{4.4}$$

3. Finally we multiply the flux of the two components by the scaling factor, i.e., parameter (d). As illustrated in Figure 4.2, the scaled combined flux (violet open circles) is the same as the predicted fluxes (black solid line) derived from the best-fit hybrids/templates.

$$f(\lambda)^{\rm total,ext}_{\rm temp} \times S = f(\lambda)^{\rm pred,ext}_{\rm temp} \tag{4.5}$$

4. After multiplying the normalized AGN and galaxy template by A and B separately, we can now measure the flux fraction of the host galaxy in the U and V bands (i.e., β_U and β_V) at effective wavelength 3778Åand 5737Å, separately. Then we apply the host fraction and the dust extinction to the observed total flux, $f^{\rm total,ext}_{\rm obs,i}$, and derived the intrinsic host flux $f^{\rm host,int}_{\rm temp,i}$ using Equation 4.6 :

$$\beta_U = \frac{f^{\rm host,int}_{\rm temp,U} \times 10^{-0.4A_U}}{f^{\rm total,ext}_{\rm temp,U}} \Rightarrow f^{\rm host,int}_{\rm obs,U} = \frac{\beta_U \times f^{\rm total,ext}_{\rm obs,U}}{10^{-0.4A_U}} \tag{4.6a}$$

$$\beta_V = \frac{f^{\rm host,int}_{\rm temp,V} \times 10^{-0.4A_V}}{f^{\rm total,ext}_{\rm temp,V}} \Rightarrow f^{\rm host,int}_{\rm obs,V} = \frac{\beta_V \times f^{\rm total,ext}_{\rm obs,V}}{10^{-0.4A_V}} \tag{4.6b}$$

5. Finally we determine the intrinsic host apparent magnitude in the U and V bands using Equation 4.7 with correction for dust extinction and AGN contribution. Based on the apparent magnitude and absolute magnitude relation ($m = M + DM$, where DM is distance modulus), we can easily obtained the corrected host absolute magnitude by Equation 4.8

$$m^{\rm host,int}_U = -2.5{\rm log}({\rm f}^{\rm total,ext}_{\rm obs,U}) - 48.59 - 2.5{\rm log}\beta_{\rm U} - {\rm A_U} \tag{4.7a}$$

$$m^{\rm host,int}_V = -2.5{\rm log}({\rm f}^{\rm total,ext}_{\rm obs,V}) - 48.59 - 2.5{\rm log}\beta_{\rm V} - {\rm A_V} \tag{4.7b}$$

$$M^{\rm host,int}_U = M^{\rm total,ext}_U - 2.5{\rm log}\beta_{\rm U} - {\rm A_U} \tag{4.8a}$$

$$M^{\rm host,int}_V = M^{\rm total,ext}_V - 2.5{\rm log}\beta_{\rm V} - {\rm A_V} \tag{4.8b}$$

Throughout this section, we take ($M^{\rm host,int}_U - M^{\rm host,int}_V$) as the dust and AGN corrected rest-frame ($U - V$) color to study the CMD for normal galaxies and different types of AGNs.

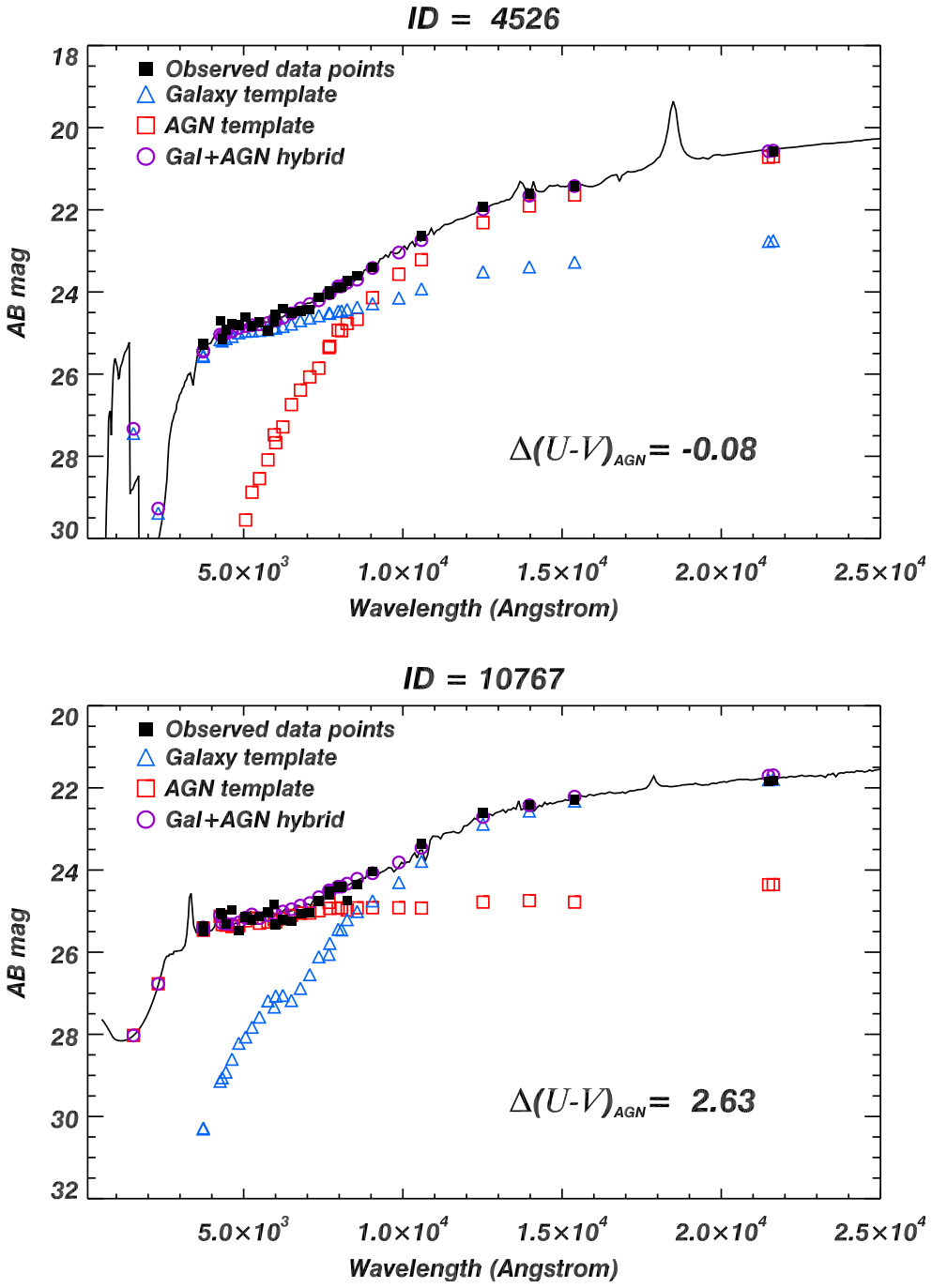


Figure 4.2: Two examples illustrating the AGN-galaxy decomposition. The upper source is best fit with hybrid "s800-2-QSO2-8" , and the lower source is best fitted with hybrid "s10-9-TQOS1-1" (black solid lines). Red squares and blue triangles indicate the decomposed contribution of AGN and galaxy to the total flux. Violet open circles are obtained summing up the two contributions. The observed data points (filled black squares) precisely lay on the top of the best-fit template, indicating that the decomposition are correctly performed. $\Delta(U-V)_{AGN}$ is the color difference due to the correction for AGN contribution. In the first case, the AGN contribution at optical bands is minimum and $\Delta(U-V)_{AGN}$ is negligible, contrary to what happens in the second case.

Moreover, based on the best-fit templates from the SED-fitting, we divide our X-ray selected AGN sample into three groups:

- AGN1: X-ray sources that are best fitted with QSO1-galaxy hybrids.
- AGN2: X-ray sources that are best fitted with QSO2-galaxy hybrids.
- Galaxy-like AGNs: X-ray sources that are best fitted with pure-galaxy templates.

In the following analysis, we will discuss in detail the color distributions of these three groups of AGNs, and compare to normal galaxies. In addition, we look at the color difference $\Delta(U-V)$ before and after all corrections described above to examine their effects. The total color differences consists of changes from the correction for dust extinction, $\Delta(U-V)_{dust}$, and changes from removing the AGN contribution, $\Delta(U-V)_{AGN}$. We will discuss these two effects separately to investigate whether the color changes are mainly due to the correction for dust or the AGN contribution.

4.2 Results

After the corrections for the dust extinction and removing AGN contribution, we investigate the color-magnitude relation for normal galaxies and AGNs, applying the criterion from Bell et al. (2004) who defined red-sequence galaxies as being redder than:

$$\langle U-V\rangle_{\rm rest,vega} = 1.15 - 0.31z - 0.08(M_{V,\rm vega} - 5\log h + 20) \tag{4.9}$$

In our work, we convert above color cut from Vega magnitude to AB magnitude using:

$$\langle U-V\rangle_{\rm rest} = -0.31z - 0.08M_{\rm V} + 0.26 \tag{4.10}$$

This color cut is 0.25 mag blueward of a linear fit to the red-sequence galaxies, and is used to divide the sample into passive galaxies (red sequence) and star-forming galaxies (green valley or blue cloud).

Figure 4.3 displays the color-magnitude relation as a function of redshift for different types of AGNs, and normal galaxies. At $z < 1.5$, we find many faint red-sequence galaxies which are rarely be found in optical selection. This can be easily understood by looking at Figure 3.12: depending on the depth of the survey, R-band selection will miss the passive objects which are fainter in the optical bands than star-forming objects. On the contrary, NIR selection will detect those passive galaxies, and depending on the depth, it will also detect an increasing number of star-forming galaxies. This makes our H-detected sample (depth down to ~ 29 mag) more complete than an optical selection in R-band (depth$\sim$24 mag) as in Bell et al. (2004). For three types of AGN hosts (i.e., template-based AGN1, AGN2, and galaxy-like AGN), they spread broadly in the CMD and have large fraction of

population placed in the green valley at all redshift bins. As redshift increases, the number of AGN host galaxies in the blue cloud region decreases and becomes absent at $z > 2.5$, with still few red-sequence sources remaining. It is interesting to find that, at high redshift, the hosts of the type 1 AGNs prefer to locate in the red sequence and have extremely red colors, while for type 2 AGN hosts, they intend to present in the green valley.

This color distribution at $z > 1.5$ is similar to the results from previous CMD works at $z \sim 1$ (e.g., Nandra et al., 2007; Silverman et al., 2008a; Cardamone et al., 2010a), which found a large fraction of AGNs in the green valley and little in the blue cloud. As described in Cardamone et al. (2010a), the AGN hosts in the red sequence have old stellar population, and their central AGN keep heating the interstellar gas and preventing the star formation in the host. The AGNs host in the green valley have young stellar population, which might be undergoing the AGN feedback of quenching star formation, leading to a migration from the blue cloud to the red sequence.

We further inspect the color distributions by measuring the distances to the color cut (defined by Equation 4.10) as a function of redshift for the normal galaxies and AGNs in Figure 4.4 and 4.5, respectively. In both figures we compare the color distributions before and after corrections for dust extinction or/and AGN contribution.

For normal galaxies, the overall color distributions become bluer after the correction for the dust extinction, which is consistent to what has been demonstrated in Brammer et al. (2009). The color bimodality appears up to $z \sim 2$, although the red peak at each redshift range is not very distinct (Figure 4.4). As we have discussed, the reason is that due to the extremely deep H-band date we used (depth $\sim$ 29 mag), a large number of faint H-band detections contributes to the blue group. This leads to the result that red distribution (positive values) are diluted by the large number of blue distribution (negative values). In addition, the blue peaks of galaxy samples at $z < 2$ are slightly bluer (~ -0.1 mag) than samples at $z > 2$. In general, the position of the blue peak stay roughly the same at value ~ -1 throughout all the redshift ranges.

For AGN host colors, after the corrections for dust extinction and AGN contribution, Figure 4.5 shows distinct color bimodality up to $z \sim 2.5$. The position of the blue peaks in the AGN samples at $z < 2$ are constant at value ~ -1, but become redder at $z > 2$. However the distributions of the red sources have more irregular behaviors. At $z < 0.5$ and $1.5 < z < 2$, they peaks in the region of red sequence (positive values), while at $0.5 < z < 1$ and $1 < z < 1.5$, the peaks blue-shifted to the region of green valley (negative values). As demonstrated in Silverman et al. (2008a), the blue-shifted red peaks at $0.5 < z < 1$ is due to a Large Scale Structure present in the field, which contributes with a large number of AGNs to the green-valley region. Finally at $2 < z < 3$, the red peaks of AGN samples become redder than samples at lower redshift. The present of bimodality we find for AGN sample, however, is contrary to the previous works who found no apparent color bimodality for AGN hosts in the same field (Silverman et al., 2008a; Cardamone et al., 2010a; Xue et al., 2010). We will discuss these discrepancies later in Section 4.4. In short, it is due to selection effects, even if those and this works are focusing on the same area.

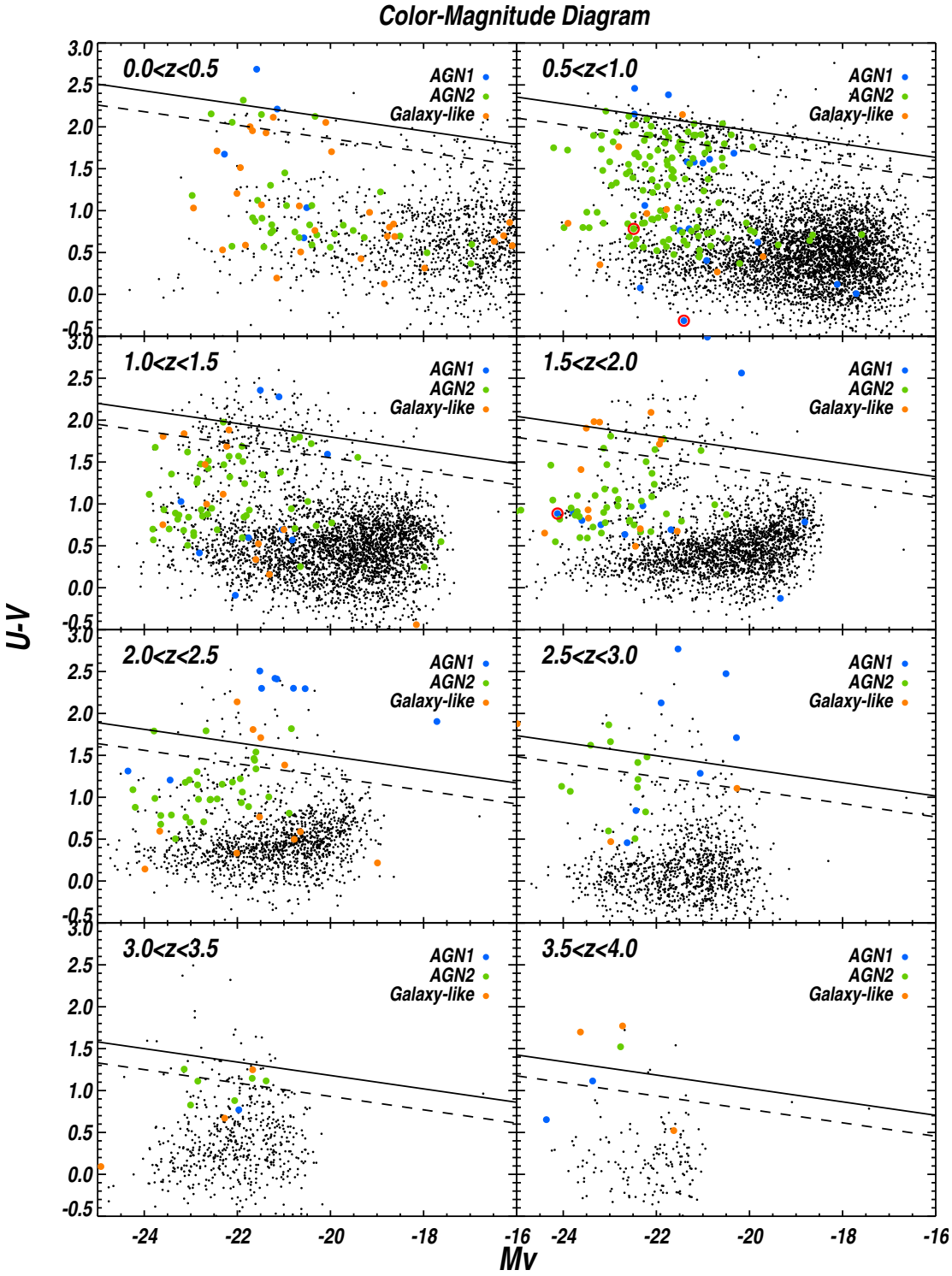


Figure 4.3: Color-magnitude diagram for normal galaxies (black dots) and AGN hosts (colored filled circle. Blue: AGN1; green: AGN2, orange: galaxy-like AGNs). Solid lines are the linear fit to the red-sequence galaxies. Dashed lines are the lower color cut for the red sequence. The red open circles represent the three sources with large $\chi^2 > 1000$ from SED-fitting results for our AGN sample, for which the composition may not be reliable.

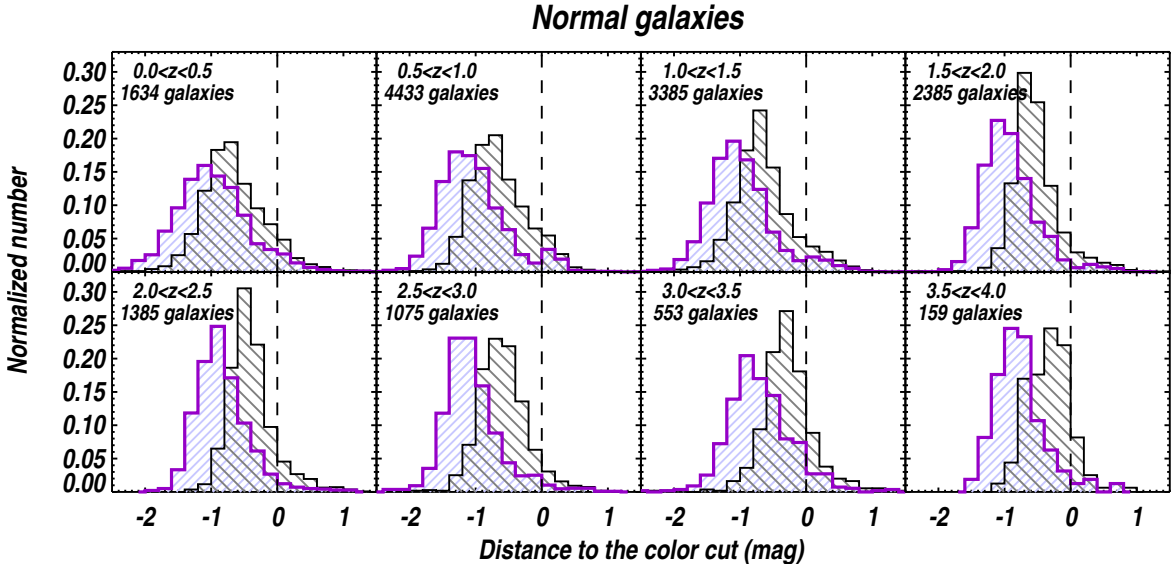


Figure 4.4: Color distributions for the normal galaxies in terms of the distances to the color cut (dashed lines, defined by Equation 4.10) before (grey histograms) and after (violet histograms) the correction for dust extinction. The color cut divides sample into passive galaxies (right) and star-forming galaxies (left).

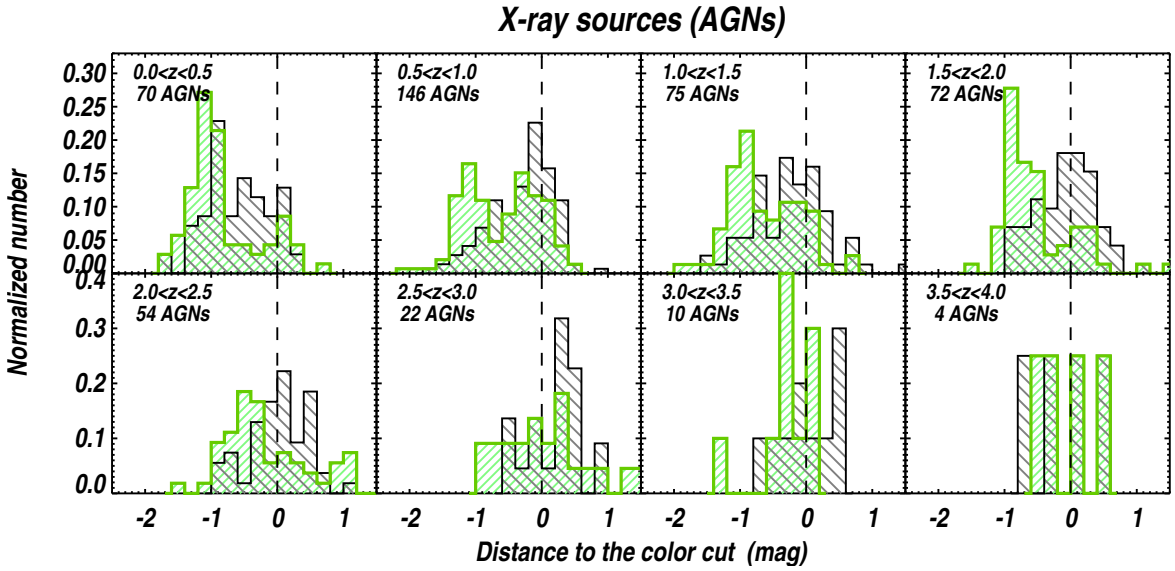


Figure 4.5: Color distributions for the AGNs in terms of the distances to the color cut (dashed lines, defined by Equation 4.10) before (grey histograms) and after (green histograms) the corrections for dust extinction and AGN contribution.

In Figure 4.6, we compare the color distributions between galaxy and AGNs after correcting for dust (for both samples) and AGN contribution. We find that: (1) in general, the red peaks of AGN hosts are bluer than galaxy sample; (2) the fractions of red groups in the AGN samples at all redshift ranges are larger than those in the galaxies; (3) the blue peaks of AGN samples at $0 < z < 1$ are consistent with galaxy samples, but becomes redder and redder with increasing redshift. To fairly compare AGNs and galaxies in the comparable mass range, we apply a absolute magnitude cut at $M_V < -20$ for both samples (i.e., mass-matched samples) and make the same comparisons again in Figure 4.7. The bimodality becomes more clear in both samples. Comparing the Figure 4.6 and Figure 4.7, the blue groups of mass-matched galaxy samples become more peaked and bluer in general. However for the AGN samples, the changes are very little as most of the AGNs are at $M_V < -20$, and we only lost few of them after applying the cut.

In Figure 4.8 and 4.9, we display the same plots as Figure 4.6 and 4.7, but without any correction either for dust extinction or for AGN contribution. Almost all the bimodalities disappear in the dust/AGN-uncorrected samples. Only at $z < 1$, luminous galaxy samples ($M_V < -20$) still present color bimodalities.

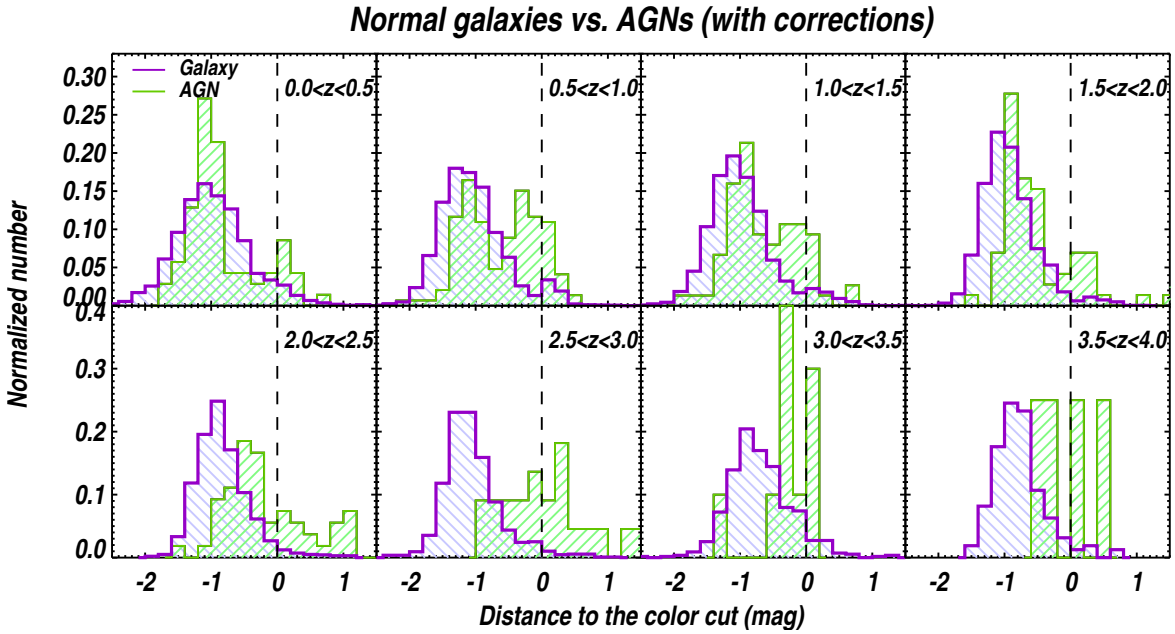


Figure 4.6: Comparison of color distributions between galaxies and AGNs with the corrections for the dust extinction or/and AGN contribution. No magnitude cut is applied on these samples.

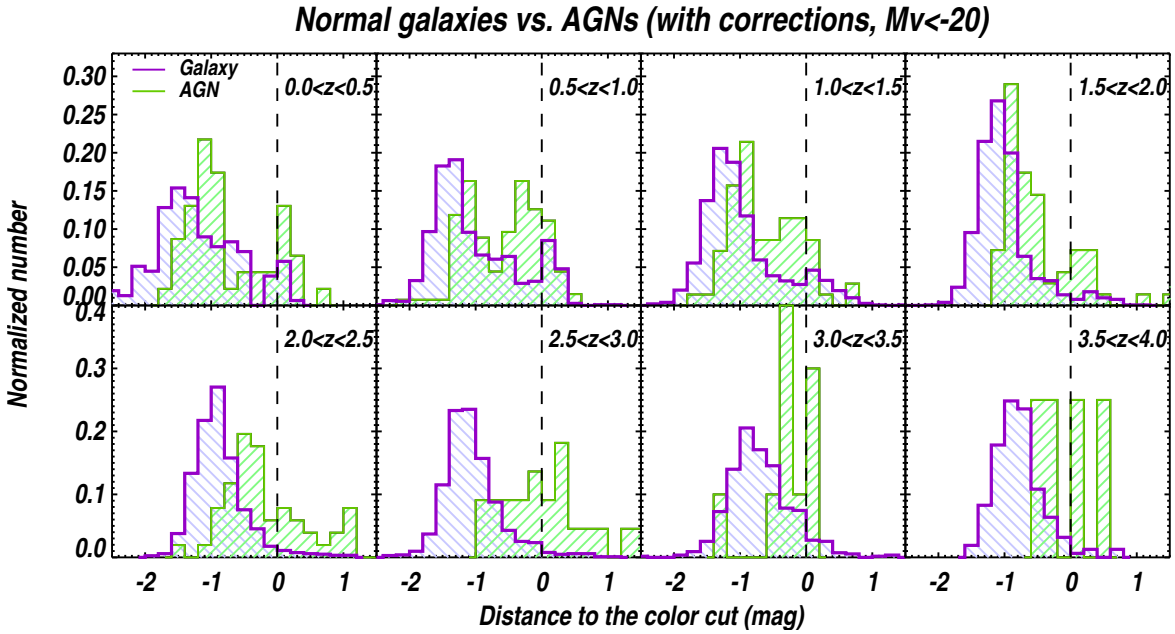


Figure 4.7: Comparison of color distributions between galaxies and AGNs with the corrections for the dust extinction or/and AGN contribution. These luminous samples are selected by the magnitude cut $M_V < 20$.

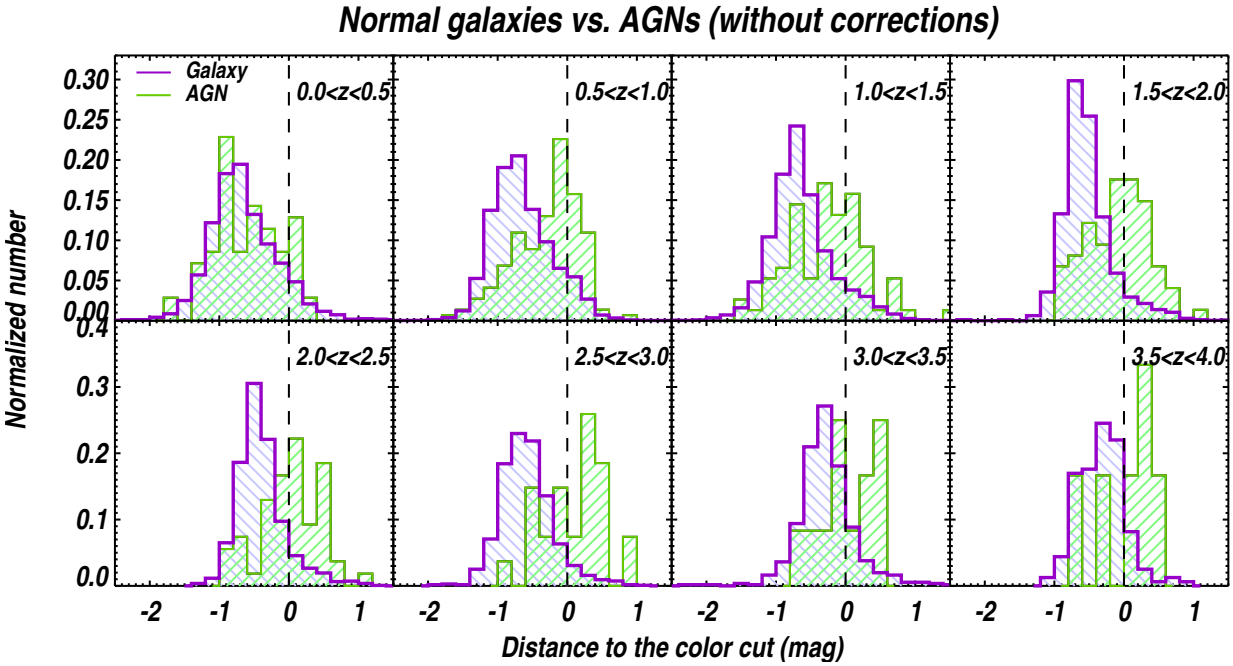


Figure 4.8: Comparison of color distributions between galaxies and AGNs without the corrections for the dust extinction or/and AGN contribution. No magnitude cut is applied on these samples.

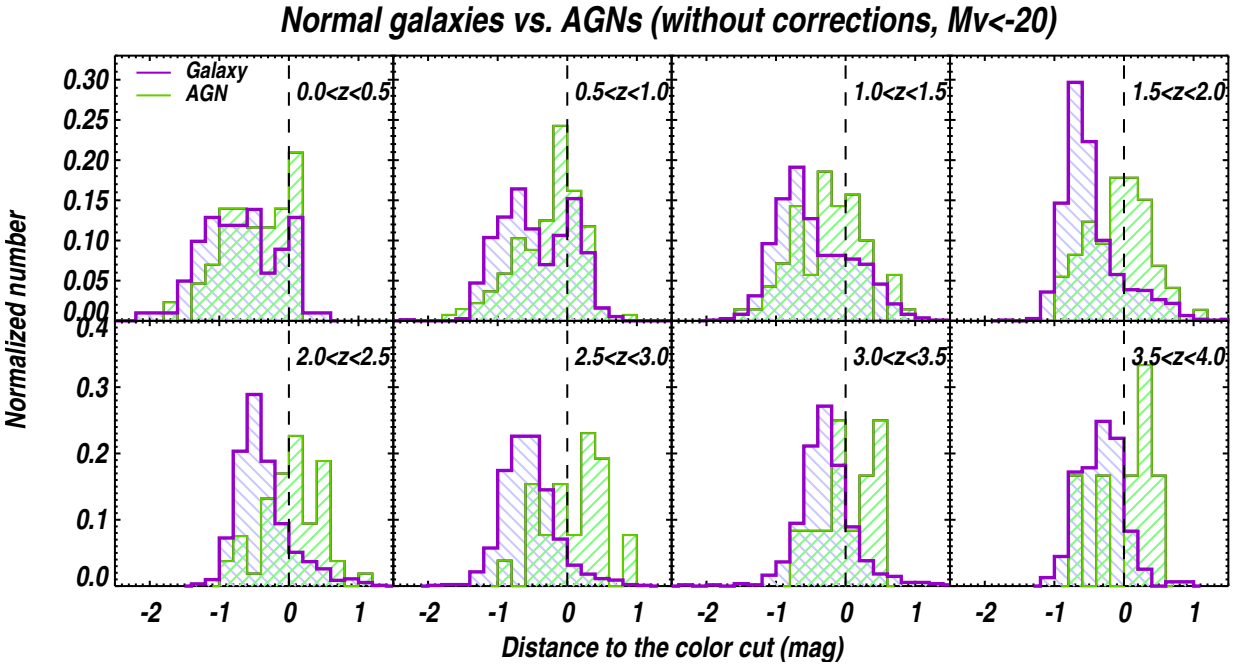


Figure 4.9: Comparison of color distributions between galaxies and AGNs without the corrections for the dust extinction or/and AGN contribution. These luminous samples are selected by the magnitude cut $M_V < 20$.

4.3 The effects of dust and AGN contribution on the host colors

In order to quantify the impact of correction for dust reddening or for AGN contribution as a function of AGN activity, we consider the color variation separately. Furthermore, we correct the dust extinction also for the normal galaxies in order to quantify weather there are differences between galaxies and AGN hosts:

- **Considering only dust correction**

  After the de-reddening, about 50% of AGN hosts moved from the red sequence to bluer region (green valley or blue cloud), which is consistent to the result from Cardamone et al. (2010a). The mean value of color changes ($\Delta(U-V)_{dust}$) for AGN is ~ -0.3. Looking detailedly in different types of AGN as a function of redshift (Figure 4.10), we find that the template-based type 2 AGNs tend to have more dust extinction that type 1 AGN. This is consistent with the picture that type 2 AGNs are more obscured than type 1. For normal galaxies, half of them also have blue de-redden colors, moving from red sequence to bluer region. The mean value of $\Delta(U-V)_{dust}$ is ~ -0.35 mag.

- **Considering only AGN correction**

  For AGN1 and AGN2, after correcting the SEDs for AGN contribution, the mean values of the color differences $\Delta(U-V)_{AGN}$ are 0.44 and -0.07, respectively (Table 4.1). This implies that AGN-dominated objects are more sensitive to the correction for the AGN host than for the de-reddening. In average, this correction seems negligible when considering a sample of AGN selected in deep X-ray observations of a small pencil beam area like the Area1 in CDFS. This is similar to the conclusion of the previous works in the same area by Silverman et al. (2008a); Xue et al. (2010) who decided to ignore the correction for AGN contribution. However, for shallower and wide areas, where the type 1 AGN become dominant (e.g. COSMOS or more dramatically, ROSAT and eROSITA), and the correction for AGN contribution to the $U-V$ color would be much more relevant than the correction for dust extinction. An interesting point shown Figure 4.10 is that while the $\Delta(U-V)_{AGN}$ is practically constant over the entire redshift range for AGN2, the variation is irregular and with a large scatter for AGN1, probably due to low number statistics.

Generally speaking, as we shown in Figure 4.10 and summarized in Table 4.1, the correction for the dust extinction makes host colors bluer, while the correction for the AGN contribution makes host colors redder.

After having understood the separate effect of correcting the $U-V$ colors for extinction and AGN contribution, it is natural to ask whether one of both corrections are responsible for moving the sources in the CMD as a function of redshift. In order to do that, we compare the color changes between the correction for dust extinction and correction for AGN

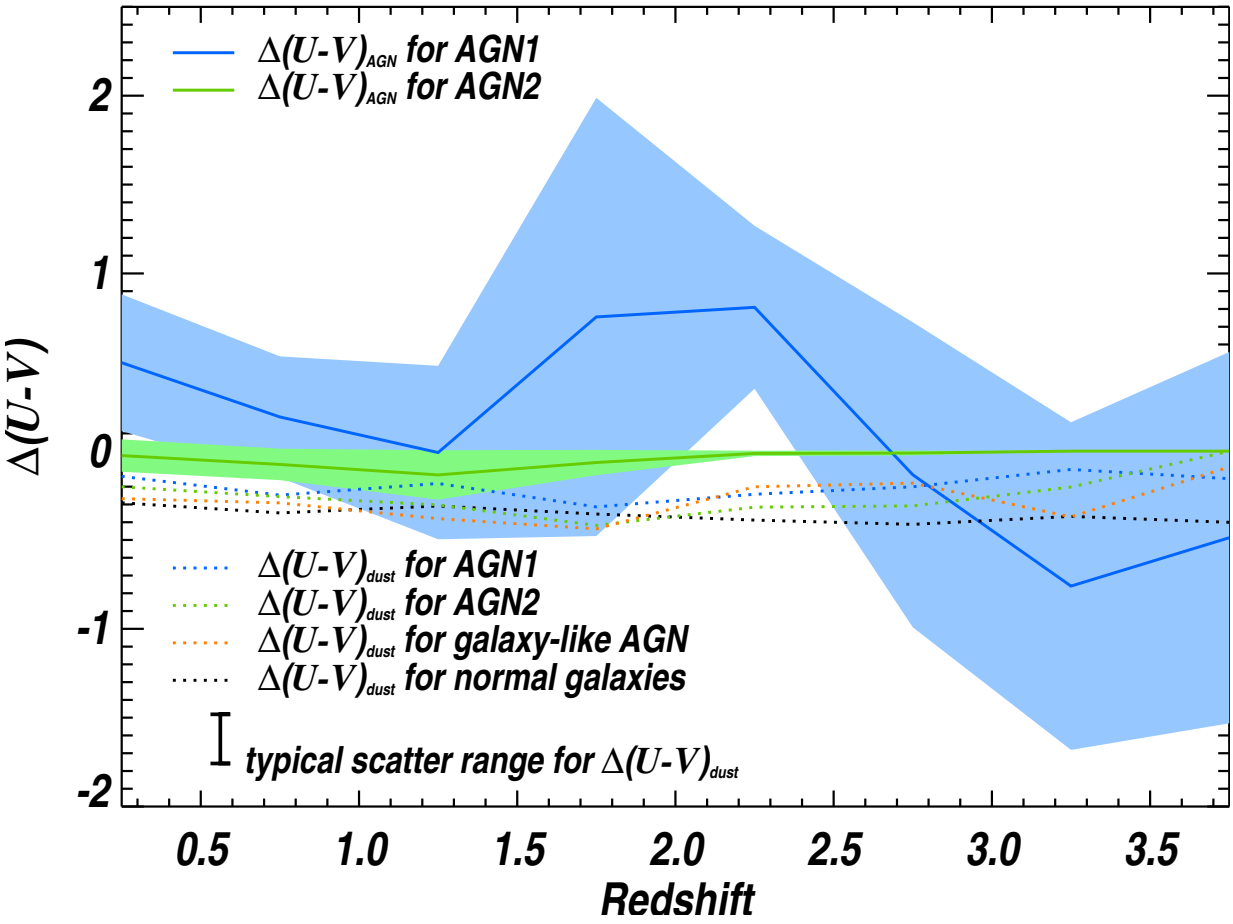


Figure 4.10: Mean values of $\Delta(U-V)$ as a function of redshift, after correcting for dust extinction (dashed lines) or AGN contribution (solid lines) for various types of sources. The shading area are the errors associated to the 1 σ scatter of $\Delta(U-V)_{AGN}$.

Table 4.1: Mean values of $\Delta(U-V)$ derived from the correction of dust and AGN contribution for AGNs and galaxies.

Correction		Dust	AGN	Dust + AGN
Non-X-ray	Normal galaxy	-0.35 ± 0.06	0.00	-0.35 ± 0.06
X-ray	AGN1	-0.31 ± 0.11	0.52 ± 0.48	0.22 ± 0.68
	AGN2	-0.30 ± 0.08	-0.07 ± 0.01	-0.37 ± 0.10
	Galaxy-like AGN	-0.29 ± 0.11	0.00	-0.29 ± 0.11

The error associated to each value is the 1 σ scatter.

contribution. The left panel of Figure 4.11 reveals that the bimodality in AGN sample is mainly caused by the correction for the dust extinction up to $z \sim 3$. The distributions of AGN-corrected host colors (the right panel) have little changes form the uncorrected

colors because of the large fraction of host-dominated AGN from the 4Ms-CDFS survey. However, again, this is due to the intrinsic nature of the sources detected in CDFS. We expect that a similar plot produced for X-ray sources detected in the XMM-COSMOS field would give a different result.

For better understanding, instead of plotting the color changes using an histogram as shown in Figure 4.11, we plot the $\Delta(U-V)$ as a function of ΔM_V in Figure 4.11. The figure reveals clearly that the corrections for dust extinction ($\Delta(U-V)_{dust}$) and AGN contribution ($\Delta(U-V)_{AGN}$) result in opposite directions of magnitude changes (ΔM_V). The former one makes sources brighter and having bluer color, while the latter leads to fainter magnitude of M_V. In the figure, we also color coded the sources as a function of their types, and compared with color changes after de-reddening for normal galaxies. The scatter of the $\Delta(U-V)_{AGN}$ in AGN1 is increasing toward higher redshift, with maximum value up to $\Delta(U-V)_{AGN} \sim 3$ at $1.5 < z < 2$. However the values of $\Delta(U-V)_{AGN}$ in AGN2 are relatively smaller than AGN1 at all redshift bins, range from -1 to 0. As we anticipated, the type 1 AGN-dominated objects have more contribution to their host colors, and are the correction for AGN contribution makes color in average redder. This can be explained by two examples in Figure 4.2 which illustrates the changes of continua slope from the hybrid (violet open circles) to pure galaxy (blue triangles) templates. The figures show that the AGN1-corrected host colors become redder (with positive $\Delta(U-V)_{AGN}$), on the contrary the AGN2-corrected host colors become bluer(with negative $\Delta(U-V)_{AGN}$).

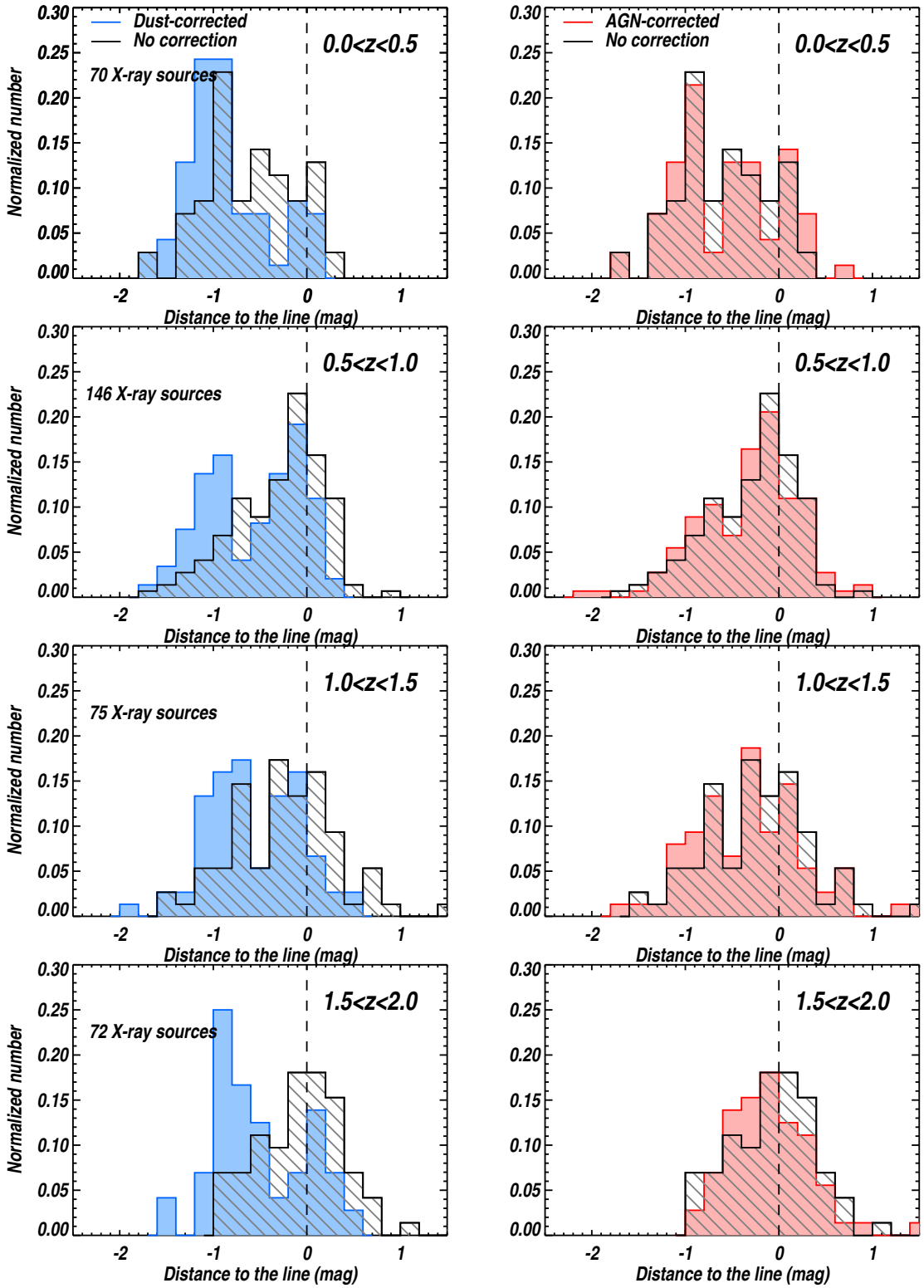


Figure 4.11: Color distributions of X-ray selected sources in terms of the distances to the color cut from Equation 4.10 before (black histograms) and after correcting for dust extinction only (blue histograms) and AGN contribution only (red histograms).

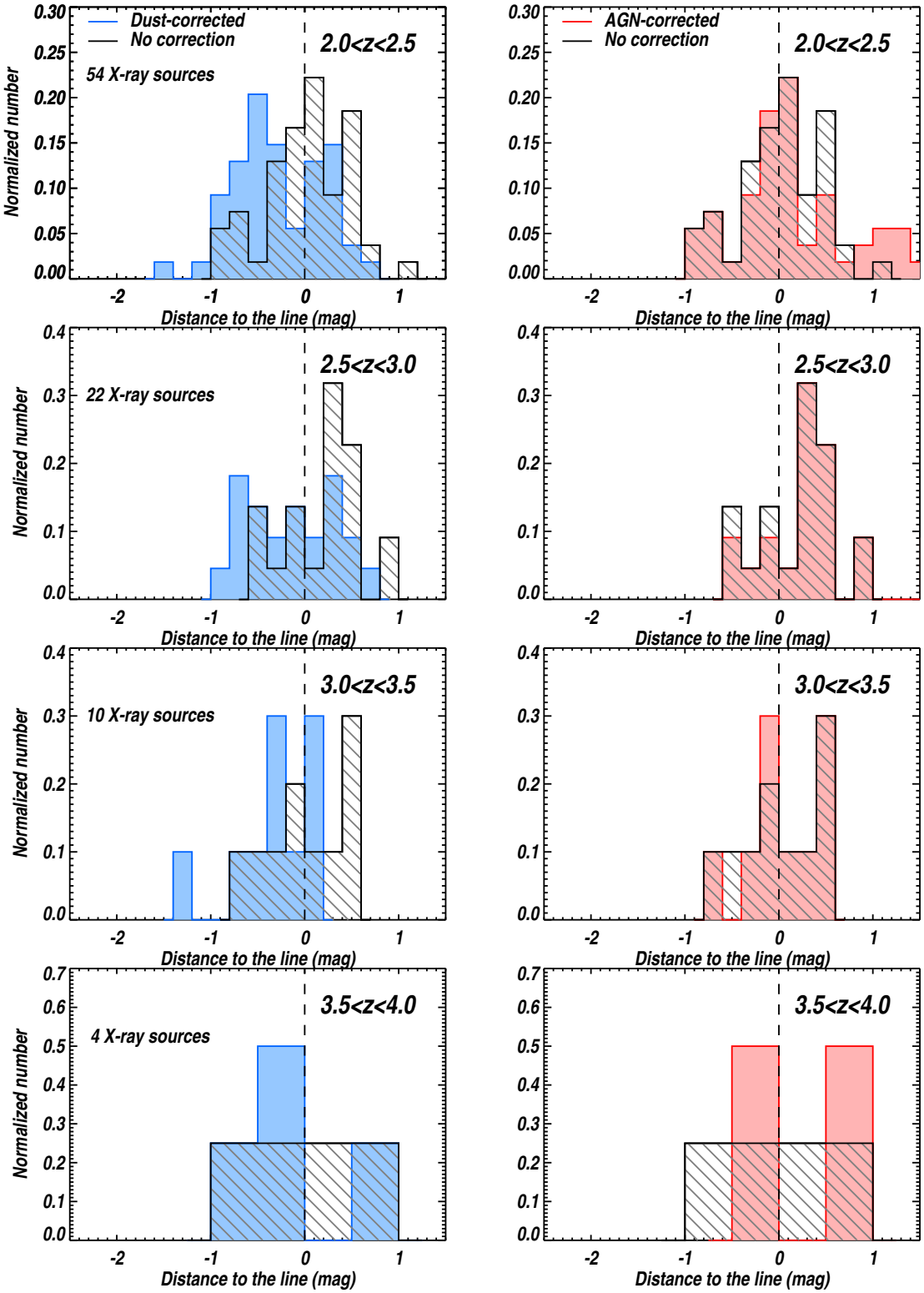


Figure 4.10: (Continued)

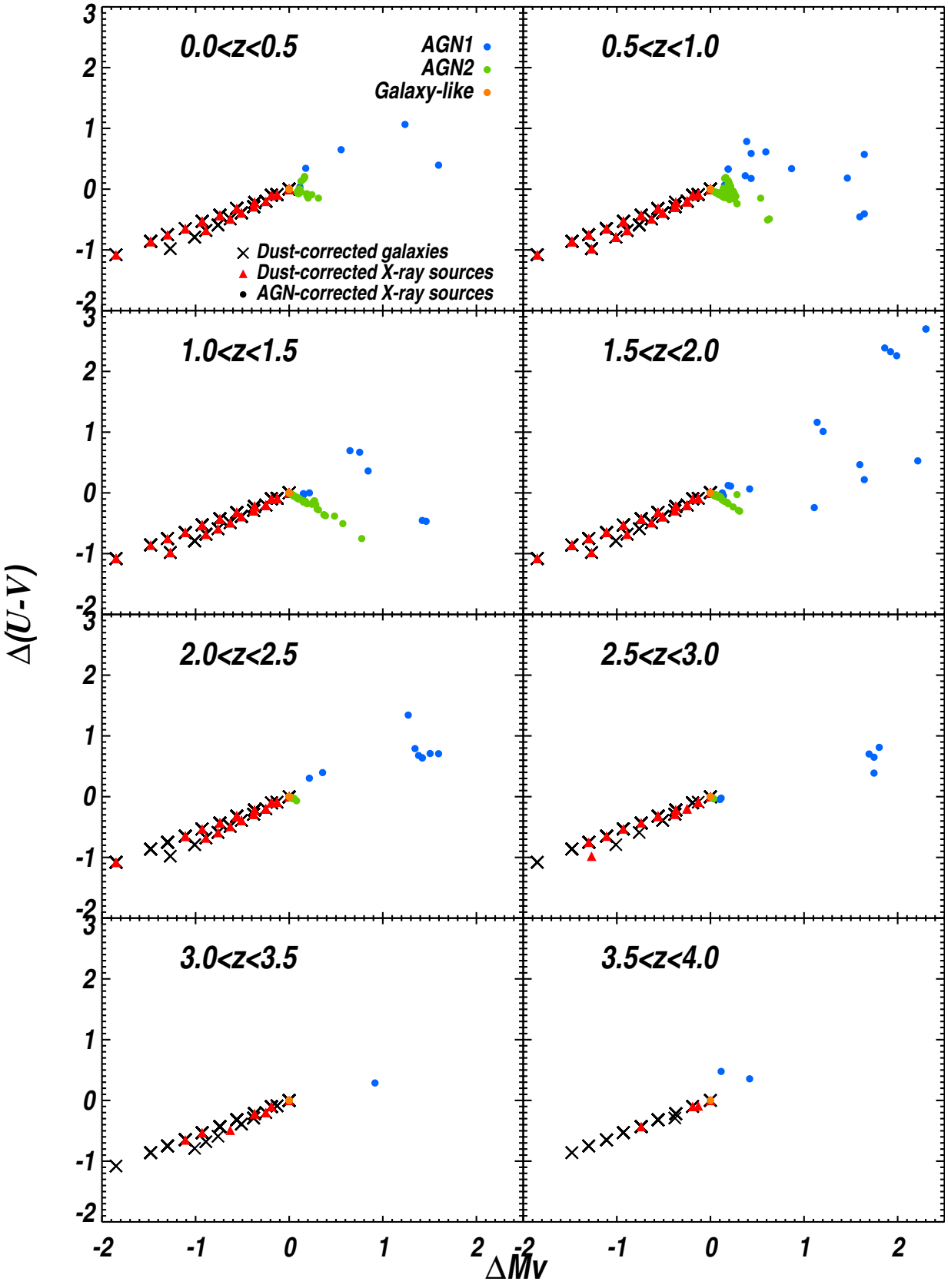


Figure 4.11: $\Delta(U-V)$ as a function of ΔM_V for the sources corrected for the dust extinction (crosses for galaxies and red triangles for AGNs) and AGN contribution (color coded dots).

4.4 Discussion

4.4.1 Comparison with previous works

• For normal galaxies:

Bell et al. (2004) studied the color-magnitude relation using R-detected galaxies in the COMBO-17 with $M_R < 24$ in the redshift range $0.2 < z < 1.1$. They found that the position of the blue peak becomes bluer toward the early Universe (with maximum color changes of ~ 0.5 mag, see Figure 1.11), which is opposite to what have been found in Williams et al. (2009) who used K-detected galaxies observed from the ground-based UK Infrared Telescope ($UKIRT$) to study CMD. In fact, they selected luminous galaxies by $K < 22.4$ at $0 < z < 2.5$ and found that the position of the blue peak becomes redder toward higher redshift (see Figure 4.12).

Our result for normal galaxies is in apparent contradiction with the results of both mentioned studies. In fact, for our H-detected (down to $H \sim 29$ mag) sample of galaxies, first we do not find a clear bimodality and secondly, the position of the blue peak keeps constant up to $z \sim 1$ and becomes slightly redder (~ 0.1 mag) at higher redshift (Figure 4.4). However, this is only due to the different depth of the samples under analysis. To illustrate this, we make the color distribution again with a magnitude cut similar to Williams et al. (2009) (i.e. $H < 22.5$) and found that with this cut the bimodality is confirmed (see Figure 4.13). Still, the position of the blue peak in our work approximately keep constant with the cosmic time. This difference can be probably explained by the fact that the sample is selected in a different band. For example, as shown in Laigle et al. 2015 (submitted) in the COSMOS field, a selection in NIR images with a depth of 24 mag misses 16% of the sources detected in i band with a depth of 26 mag. Therefore in our work, with H band selected sources down to a magnitude of ~ 29, we have more complete samples than those usually adopted for this kind of analysis.

• For AGNs:

After correcting for dust extinction and AGN contribution, we did find color bimodality up to $z \sim 2.5$ which is somehow in contrast with the previous work done in the same field by Silverman et al. (2008a) who did not find it. In Silverman et al. (2008a), they included only galaxy-like sources and excluded luminous type 1 AGN, and assumed that the optical emission of their sample are dominated by star light from the hosts. In particular, they only used spectroscopic sources to have more confident redshift, so that their analysis is restricted to lower redshift ($0.4 < z < 1.1$). Unlike Silverman et al. (2008a), Cardamone et al. (2010a) have shown few AGN hosts in the green valley, and found a bimodal distribution of X-ray selected sources in the redshift range $0.8 < z < 1.2$ after correcting the data for dust extinction. However in that specific redshift range, despite the bimodality, we find a large fraction of sources in the green valley. The difference between our work and Cardamone et al. (2010a) is probably due to the templates in which the SEDs are fitted. Cardamone et al. (2010a) only use a power-law type 1 AGN added to various galaxy templates to fit all the AGN SEDs. Therefore the decomposition in the two contributions

might be effected. Nevertheless in our work, we use well-trained AGN1/AGN2-galaxy hybrids to obtained more accurate photo-z and SED decomposition.

4.4.2 Global evolution of AGN host color

The changes of galaxy and AGN host colors with redshift is generally interpreted as evolution of stellar population and dust contents with cosmic time. In our work, since the position of the blue peaks for both galaxies and AGN host colors are almost constant, we can not claim any evolution of star formation activity in the AGN host galaxies. We already discussed how the selection of the sources can create biases in the data analysis. This is not only true for galaxies, but also for AGNs. For example, at any redshift, we find more red AGN hosts than normal galaxies, which suggests that we are biassed toward bright sources. In fact, at a given depth, a source would be more likely detected if it hosts an AGN, as the additional AGN contribution to the luminosity would make the source brighter. Another factor that makes our result different from literature is that we use absolute magnitude in the X axes rather than stellar masses. This is because the stellar masses of the AGNs have huge uncertainties due to the degeneracies in the fits, as shown in the work of Bongiorno et al. (2012) and Santini et al. (2012). This limits the use of the mass measurement and in particular does not allow the construction of an unbiassed control sample of normal galaxies matched in masses. Using the absolute magnitude as a proxy for masses and creating a sample of galaxies and AGN with $M_V < -20$, we find that the position of the blue peak for galaxies is systematically bluer than the blue peak in the AGN host distribution at any redshift.

As Xue et al. (2010) have demonstrated, we confirm that the majority of the sources detected in the 4Ms-CDFS area is composed by type 2/obscured AGN. Most of them are located in the green valley throughout all redshift ranges, implying that type 2 AGNs usually reside in younger, star-forming host galaxies. In our work, the green-valley sources have no obvious trend with redshift, and we didn't find any clear evident shows that this transition area refers to the evolutionary stage between star-forming to passive galaxies.

For the few Type 1 AGNs in this field, we find that most of their hosts are extremely red and massive, which indicates that Type 1 AGNs prefer reside in old massive early-type galaxies. However, for a more complete picture we will need to assemble a larger sample of bright AGN typical of shallower X-ray survey (e.g., XMM-COSMOS) and apply the same decomposition as in this work.

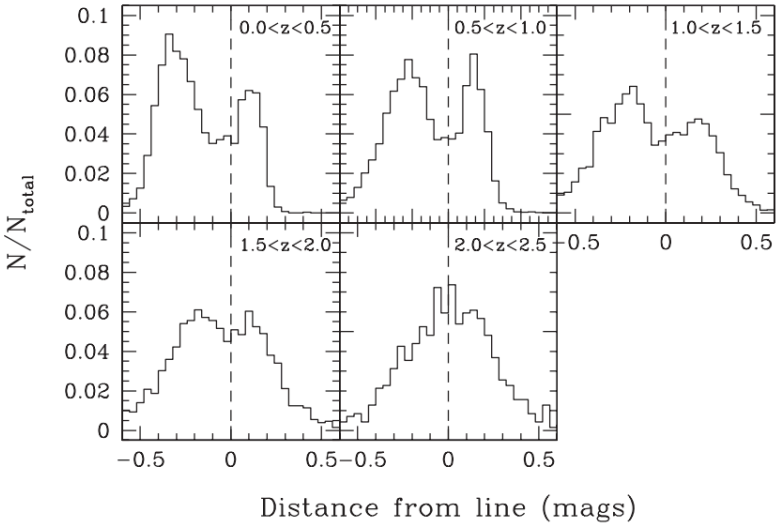


Figure 4.12: Color distributions for the K-selected galaxies in terms of the distances to the color cut from the work of Williams et al. (2009). Figure taken from Williams et al. (2009).

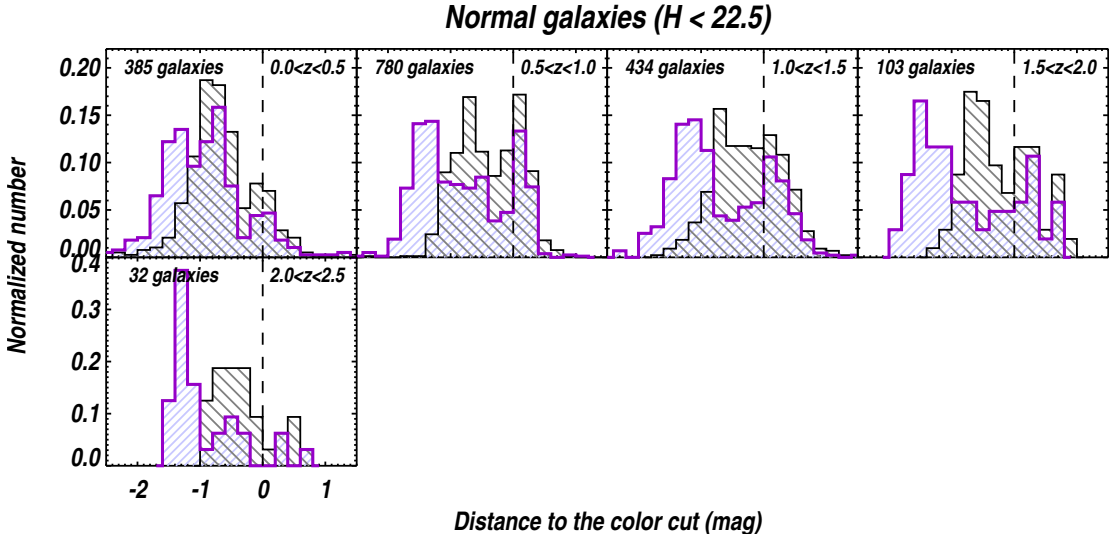


Figure 4.13: Color distributions for the normal galaxies selected by $H < 22.5$ in terms of the distances to the color cut (dashed lines, defined by Equation 4.10) before (grey histograms) and after (violet histograms) the correction for dust extinction.

4.5 Summary

In this section, we have investigated the color-magnitude diagram for normal galaxies and AGNs in the CANDELS/ GOODS-S region (i.e., Area 1), before and after correcting for extinction or/and AGN contribution. We have found that:

- As found in normal galaxies, we found that AGN host colors also present bimodality in the CMD up to $z \sim 2.5$, after properly taking into account and removing dust and AGN contributions. This implies that there are no fundamental differences between a normal galaxy and one hosting an active SMBH.

- Comparing"mass-matched" ($M_V < -20$) AGN and galaxy samples, AGN hosts have older (redder) blue group and younger (bluer) red group than galaxies. The peaks of blue group in both AGN and galaxy samples are roughly constant at $z < 2$ either with or without the absolute magnitude. As we discussed this can be due to the very deep data used for the constructions of the samples.

- We found that most of the type 1 AGN hosts appear in the red sequence, while type 2 AGN hosts tend to appear in the blue cloud or green valley. However, we didn't find any trend of evolution showing that the green-valley sources are in the evolutionary stage between star-forming and passive galaxies.

- We confirm that, at least in CDFS, the bimodality of AGN host colors in CMD is the mainly the result of the correction for the dust extinction rather than the correction for the AGN contribution. However, the few bright AGNs in this field suggest that this is an observational result that in shallower and wide areas where this type of objects will be more frequent, and we will probably reach a different conclusion. We plan to make the same analysis for the XMM-COSMOS sources to verify this statement.

Chapter 5

High-redshift AGN

To understand how SMBH form and grow and how they related to galaxies in the early Universe, redshift is the most basic parameter needed to be determined first. Because spectroscopy is difficult at high redshift, samples of galaxies at redshift $z > 2$ have in the past been selected using characteristic features in the SED that would be revealed by combinations of different colors. This, for example, is the case for the Lyman break galaxy (LBG) technique (Steidel et al., 2003) and BzK color-color selection (Daddi et al., 2004). These two methods were originally tuned for galaxies, used custom filters and were adopted on very deep images. Now, thanks to advances in technology that permit deep and detailed observation in NIR/MIR, we can better define the sources at redshift $z \sim 2-3$ and even attempt to extend the search at higher redshift by trying to shift the colors used for the selection, to longer wavelength. It is unclear however, how pure and complete these selections are, how they are effected by the chosen filter, whether they can be used for selecting AGN as well as normal galaxies, and whether a method is better than the other.

In chapters 2 and 3 we have already described how we succeeded in computing reliable redshifts for faint and distant galaxies. We can then use photo-z to answer the previous questions, not only by looking at the best-fit photo-z value but taking into account the entire redshift probability distribution function $P(z)$. In detail, we can measure the completeness and purity of color-selected samples using the redshift weighting factor w (Aird et al., 2010).

Finally, using the $P(z)$ technique, we provide a high-z source list at $z > 3$ and compared our list with previous work, demonstrating the reasons that cause the discrepancies. This high-z source list can be utilized in the future studies of the luminosity function or mass function to understand the AGN and galaxy (co)evolution in the distant Universe.

5.1 Color selection techniques for galaxies and AGNs

When photo-z (which take into account the entire shape of SED for determining the redshifts) are not available, simple color selection techniques (which use only two or three photometric points) have been historically used for selecting high redshift sources. In this

section, we investigate the capability of the LBG and BzK color techniques for selecting high-z samples of X-ray-selected sources, compare to accurate photo-z.

5.1.1 Lyman break galaxies (LBG)

The Lyman break technique is a robust broad-band color selection tuned to find high-z galaxies. This method uses the distinguishable feature of the "break" in the UV/optical continuum which is caused by the hydrogen absorption related to (1) O and B stars in star-forming galaxies, or (2) intergalactic medium (IGM) surrounding AGNs. The absorption results in a sharp drop at the rest-frame Lyman limit (912 Å), which is shifted into the visible wavelength at redshift $z \sim 3$.

Steidel et al. (1996) used this method to search for Lyman break galaxies (hereafter $\mathrm{LBG_{S96}}$) at $z \sim 3$ using ground-based imaging in broad filters (specifically U_n, G, R) as listed in the Table 1.1. Since then, the Lyman break technique has been extended to the search for sources at $z > 3$ and has also been used to define samples of AGNs at high redshift (e.g., the studies of X-ray space density and luminosity function in Nandra et al., 2005; Laird et al., 2005; Lehmer et al., 2005b; Aird et al., 2008, 2010).

In our work, since we used space-based HST filters which are not the same filters used in Steidel et al. (1996), we can not directly use the original criteria listed in Table 1.1. Instead we adopted the filter-calibrated color criteria for LBG selection defined by Vanzella et al. (2009) (hereafter $\mathrm{LBG_{V09}}$) to select sources at $z \sim 3 - 4.5$:

$(B_{435} - V_{606}) > 1.1 + (V_{606} - z_{850}) \wedge$
$(B_{435} - V_{606}) > 1.1 \wedge$
$(V_{606} - z_{850}) < 1.6$

These criteria are revised based on the criteria fine-tuned by Giavalisco et al. (2004b) to the GOODS/ACS survey, slightly extending the selection cut $(B_{435} - V_{606})$ bluer and $(V_{606} - z_{850})$ redder. Papovich et al. (2004) also applied these modified criteria to select the LBG (hereafter $\mathrm{LBG_{P04}}$) in the same field. We will compare their works with our result in the later section.

As we show in Figure 5.1, there are 31 X-ray sources that are selected by this $\mathrm{LBG_{V09}}$ method and their detailed information, including redshift, are listed in Table 5.1. It appears that a large fraction of the sources is not in the expected redshift range (i.e., $3.4 < z < 4.5$ for Papovich et al. 2004 and $3.1 < z < 4.4$ for Vanzella et al. 2009). This can also be seen in the bottom histogram of Figure 5.1. A similar redshift distribution was found in Dahlen et al. (2010). 11 of 31 $\mathrm{LBG_{V09}}$-selected AGNs have $z < 2.5$. For these 11 low-redshift interlopers, 7 are spectroscopically confirmed and the photo-z we obtained are fully consistent with the spec-z. Thus we can rule out the possibility of the catastrophic failure from the photo-z computation. It is likely that the misclassification of $\mathrm{LBG_{V09}}$ is due to the misinterpretation of the Lyman break as Balmer breaks as discussed in Dahlen et al. (2010).

Table 5.1: Information from the SED fitting of LBG_{V09}-selected sample in Area 1

ID_{H14} (1)	*zbest* (2)	$w(3.1 < z < 4.4)^a$ (3)	$w(3.3 < z < 3.8)^b$ (4)	mod (5)	$Name_{mod}$ (6)
24636	3.7	0.87	0.7	13	mod-fdf4.sed
6294	3.7	1.0	1.0	108	s010-3-TQSO1-7.sed
5479	3.66*	1.0	1.0	102	s010-9-TQSO1-1.sed
6905	3.639	0.99	0.97	130	s670-9-TQSO1-1.sed
5367	3.519	0.71	0.63	16	mod-s700.sed
8833	3.515	1.0	0.99	47	s670-6-QSO2-4.sed
9698	3.489	1.0	0.98	48	s670-9-QSO2-1.sed
3372	3.47*	1.0	1.0	19	sac-5-TQSO1-5.sed
12309	3.438	0.87	0.7	45	s230-8-QSO2-2.sed
19505	3.417	0.99	0.91	121	fdf4-9-QSO2-1.sed
14587	3.193*	1.0	0.0	127	s800-2-TQSO1-8.sed
5064	3.05	0.44	0.21	36	s200-9-QSO2-1.sed
7061	2.984	0.41	0.17	40	s400-9-QSO2-1.sed
5827	2.973	0.0	0.0	102	s010-9-TQSO1-1.sed
1381	2.688	0.0	0.0	37	s250-8-QSO2-2.sed
135	2.643	0.01	0.0	20	s020-9-TQSO1-1.sed
3943	2.608	0.01	0.0	4	mod-s020.sed
2408	2.597	0.01	0.0	103	s020-9-TQSO1-1.sed
9089	2.577	0.0	0.0	115	s090-6-QSO2-4.sed
3776	2.562*	0.0	0.0	18	ec-6-TQSO1-4.sed
18547	2.107	0.0	0.0	15	mod-s670.sed
4634	1.944	0.0	0.0	41	s500-8-QSO2-2.sed
12466	1.625*	0.0	0.0	121	fdf4-9-QSO2-1.sed
17713	1.37*	0.0	0.0	120	s800-5-QSO2-5.sed
17269	1.114*	0.0	0.0	36	s200-9-QSO2-1.sed
22173	1.041*	0.0	0.0	107	sac-2-TQSO1-8.sed
18694	0.608	0.03	0.01	25	s800-8-TQSO1-2.sed
14465	0.538*	0.0	0.0	1	mod-e.sed
18969	0.502	0.13	0.13	25	s800-8-TQSO1-2.sed
24012	0.152*	0.0	0.0	9	mod-s150.sed
26649	0.103*	0.0	0.0	4	mod-s020.sed

Sources with spec-z are marked by $*$.

[a] Target redshift range defined by Vanzella et al. (2009)

[b] Redshift range where the LBG_{V09}-selected AGNs has the maximum $(C + P_s)$.

(1)ID_{H14}: ID from Hsu et al. (2014)

(2) zbest is the spec-z (marked with $*$), otherwise photo-z.

(3) photo-z weight at $3.1 < z < 4.4$.

(4) photo-z weight at $3.3 < z < 3.8$.

(5) mod refers to the sequence number of the template.

(6) $Name_{mod}$ is the name of the template (see detailed explanation of the templates in Bender et al. (2001) and Section 3.2).

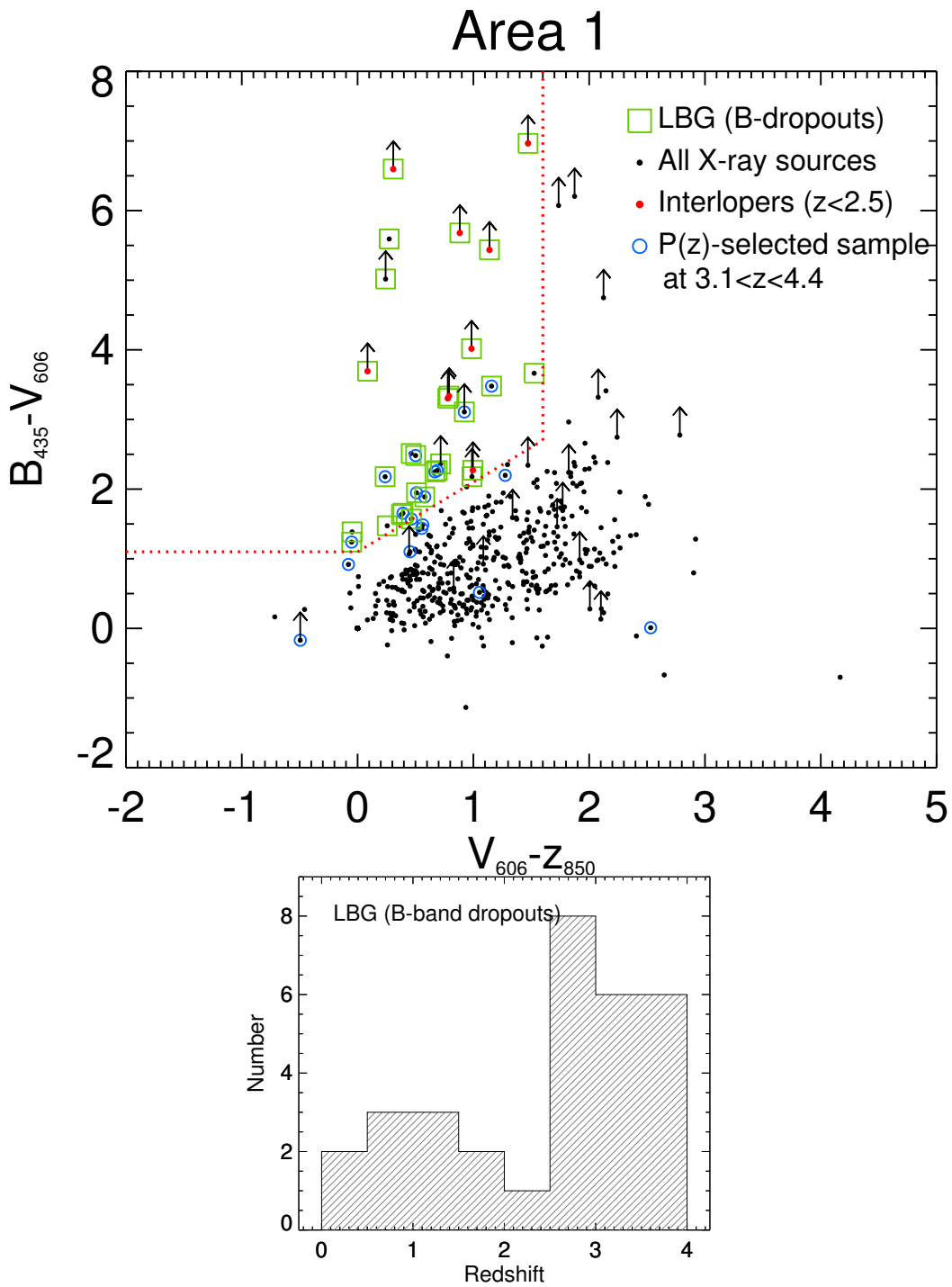


Figure 5.1: Top: $\mathrm{LBG_{V09}}$-selected AGNs using the B_{450}-band dropout technique (red dotted line) in the GOODS-S (Area 1). The filled circles are all the X-ray sources in this area, while the low-redshift interlopers are marked in red; green squares are the $\mathrm{LBG_{V09}}$-selected AGNs. The arrows refer to the lower limit of the color $(B_{450} - V_{606})$. Blue open circles indicate the $P(z)$-selected sample at $3.1 < z < 4.4$ which will be discussed in the later section. Bottom: Redshift distribution for the $\mathrm{LBG_{V09}}$-selected sample. The redshifts in this plot are best-fit photo-z from our work or spec-z, when available.

5.1.2 Bzk-like color-color selection

BzK color-color selection is an efficient method of selecting galaxies at $z \sim 2$ using the feature of the Balmer break (4000 Å) which is caused by dust absorption in star-forming regions. The method was proposed by Daddi et al. (2004) who have successfully separated and identified star-forming and passive galaxies at $1.4 < z < 2.5$. With deeper NIR data from HST/WFC3, we can extend this color selection to higher redshift, by adopting the criteria presented in Guo et al. (2013) as follows:

- $1.4 < z < 2.5$: $BzK = (z - K_s) - (B - z)$, where B and z are F435W and F850LP, respectively. Star-forming galaxies are selected by $BzK > -0.2$, while passive galaxies are selected by $BzK \leq -0.2$ and $(z - K_s) > 2.0$
- $2.5 \leq z < 3.5$: $VJL = (J - L) - 1.2 \times (V - J)$, where V, J, L are F606W, F125W, and IRAC 3.6 μm, respectively. Star-forming galaxies are selected by $VJL > 0$, while passive galaxies are selected by $VJL \leq 0$ and $(J - L) > 2.5$
- $3.5 \leq z < 4.5$: $iHM = (H - M) - 1.375 \times (i - H)$ where i, H, M are F775W, F160W, and IRAC 4.5 μm, respectively. Star-forming galaxies are selected by $iHM > 0$, while passive galaxies are selected by $iHM \leq 0$ and $(H - M) > 2.0$

Applying the criteria above, we found 177, 114 and 88 high-z AGNs selected by the BzK, VJL and iHM methods, respectively (see Figure 5.2, 5.3 and 5.4). The right panels of Figures 5.2, 5.3 and 5.4 show that the redshift distributions of these samples are generally broad, and strongly contaminated by low-redshift interlopers as in the case of the $\mathrm{LBG_{V09}}$. However for BzK and VJL samples, they peaked roughly in the expected ranges, having the median value of $z \sim 1.9$ and $z \sim 2.4$, respectively.

5.1.3 Results

In the left panels of Figure 5.1, 5.2, 5.3, and 5.4, low redshift interlopers are indicated. It can be see that the four criteria are increasingly unreliable for selecting X-ray sources at higher redshift, ranging from a success rate of $\sim 65\%$ for the $\mathrm{LBG_{V09}}$ to $\sim 10\%$ for the iHM. Although supported by spec-z in some cases, the analysis can still be affected by the uncertainty of the best-fit values of photo-z. In Section 5.4, we will repeat the analysis taking into account the entire $P(z)$.

5.2 High-z selection via *zbest*

When spec-z or best-fit photo-z (together called *zbest* in this thesis) are available, one can select sources within certain ranges using these pure values. However at high redshift, where spectroscopic redshifts are rare, relying on the best-fit values of photo-z can be very dangerous, especially when only few photometric points with large errors are used for SED fitting. For example, Fiore et al. (2012) defined a high-z AGN candidates in the

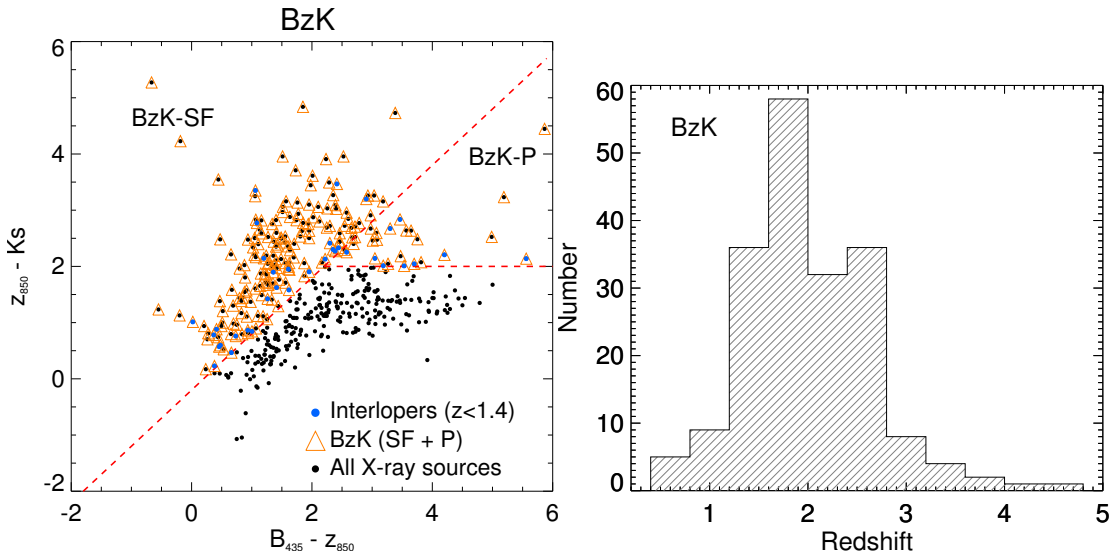


Figure 5.2: Left: BzK-selected AGNs in GOODS-S (Area 1). Red dashed lines indicate the color criteria; orange triangles represent the sources selected by the method; the filled circles indicate all the X-ray sources, while the low-redshift interlopers are marked in blue. Right: Redshift distribution for BzK selected sources. The redshifts applied in this plot are photo-z from your work or spec-z, when available.

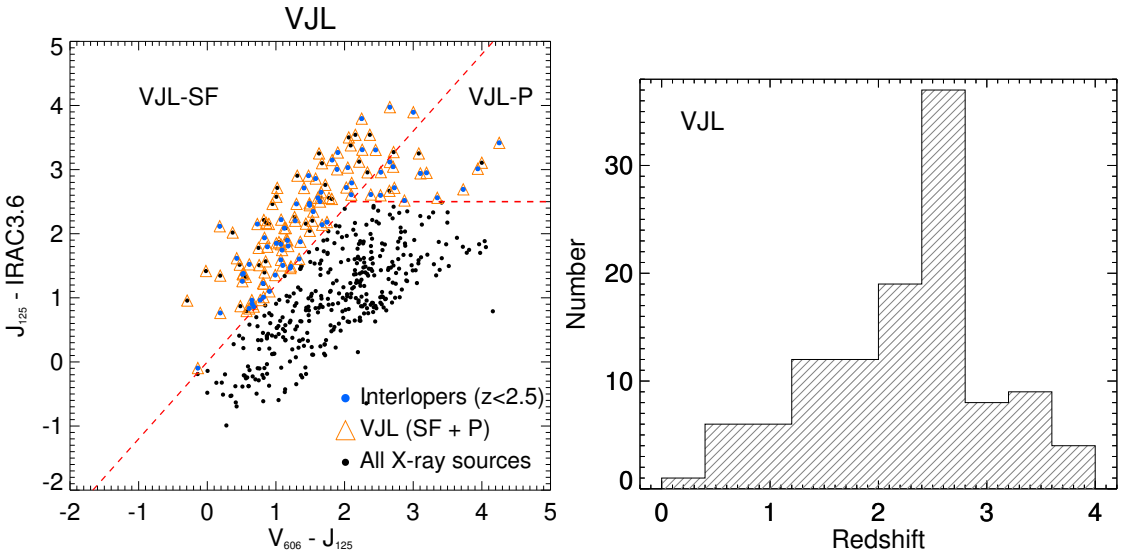


Figure 5.3: Left: VJL-selected AGNs in GOODS-S (Area 1). Red dashed lines indicate the color criteria; orange triangles represent the sources selected by the method; the filled circles indicate all the X-ray sources, while the low-redshift interlopers are marked in blue. Right: Redshift distribution for VJL selected sources.

GOODS-S using *zbest* values with the best-fit photo-z fitted by normal galaxy templates. We will compare their result with our $P(z)$-selected sample, and make further investigation

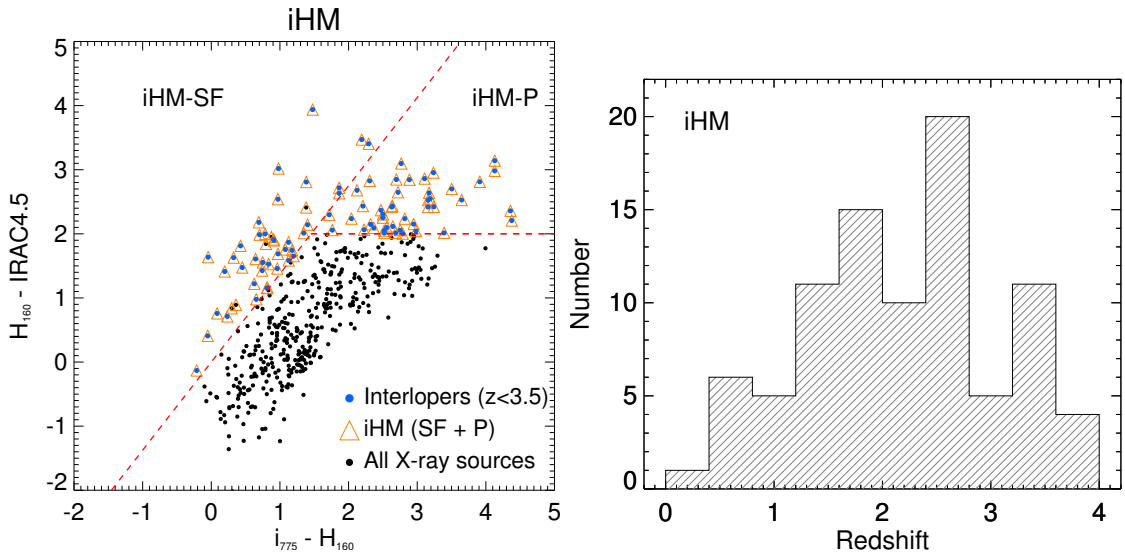


Figure 5.4: Left: iHM-selected AGNs in GOODS-S (Area 1). Red dashed lines indicate the color criteria; orange triangles represent the sources selected by the method; the filled circles indicate all the X-ray sources, while the low-redshift interlopers are marked in blue. Right: Redshift distribution for VJL selected sources.

in Section 5.5.

5.3 High-z selection via $P(z)$

In the previous sections we defined the "interlopers" as those sources that did not have photo-z or spec-z within the targeted range. However, redshifts have an error associated with them and in particular photo-z carry information in the $P(z)$ on the entire redshift range from 0 to 7. Using this information, we can further investigate the efficiency of the $\mathrm{LBG}_{\mathrm{V09}}$, BzK, VJL, and iHM methods by studying the purity (P_s) and completeness (C) of samples selected using these methods. To do this, we introduce the concept of the weight factor w (e.g., Aird et al., 2010). The value is obtained by measuring the fraction of area under the curve of $P(z)$ at a given redshift range ($z_{\mathrm{low}} < z < z_{\mathrm{up}}$), which reflects the confidence that a given source falls within a target redshift range (as illustrated in Figure 5.5). For each individual source with photo-z, we estimated the value of w as determined in Equation 5.1. [1]

$$w = \frac{\int_{z_{\mathrm{low}}}^{z_{\mathrm{up}}} P(z)dz}{\int_{0}^{7} P(z)dz} \tag{5.1}$$

[1]Note that theoretically the integration of $P(z)$ should be from 0 to ∞. However we integrate from 0 to 7 which is the range of redshift obtained from our photo-z computation with code *LePhare*.

For the sources with reliable spec-z [2], we assigned a weight of $w = 1$ if the spec-z falls within the target redshift range, while we set $w = 0$ if it is outside the target range.

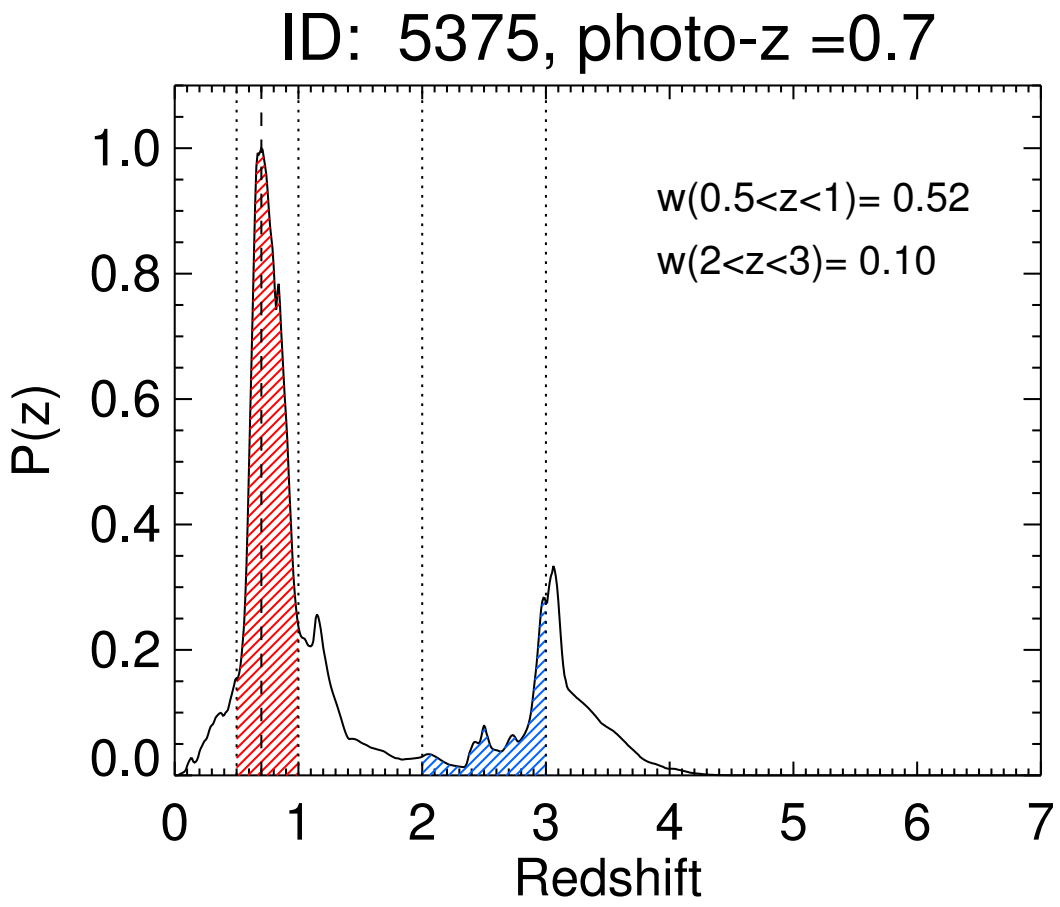


Figure 5.5: Example of measuring the value of weight w from the $P(z)$. The fraction of red shaded area ($w = 0.52$) represents the photo-z weight in the redshift range $0.5 < z < 1$. The fraction of blue shaded area ($w = 0.1$) represents the photo-z weight in the redshift range $2 < z < 3$. The dashed line is the best solution of the photo-z, $z_p = 0.7$.

We use w to estimate the C and P_s for each of the methods introduced in the previous section and verify which is the most efficient method, i.e., the sample with the highest $(C + P_s)$. We define the completeness as:

$$C = \frac{\sum_{j=1}^{N_{\rm sample}} w_j}{\sum_{i=1}^{N_{\rm total}} w_i} \tag{5.2}$$

[2]Here we select reliable spec-z by the criterion of spec-z quality $Q_{\rm sz} \leq 2$, see the detailed explanation in Table A.4

And the purity as:

$$P_s = 1 - \frac{\sum_{j=1}^{N_{\rm sample}} (1 - w_j)}{N_{\rm sample}} \tag{5.3}$$

$N_{\rm total}$ is the total number of entire sample, and $\sum w_i$ gives an effective total number of the sample. $N_{\rm sample}$ is the total number of the high-z sample selected by a given threshold of weight ($w_{\rm th}$) at a certain redshift range $z_{\rm low} < z < z_{\rm up}$, and $\sum w_j$ is an effective total number of the high-z sample.

Figure 5.6 shows how the values of C and P_s change with $w_{\rm th}$. With larger $w_{\rm th}$, less sources are selected among the entire sample, which leads to a lower completeness (C). In contrast, the purity (P_s) becomes higher with higher $w_{\rm th}$, which indicates that the fraction of interlopers is decreasing. The goal is to find the value of w that maximizes the sum of C and P_s. We do that by calculating the values of $(C + P_s)$ as a function of w with a step of 0.05 and find the maximum value of $(C + P_s)$ that fall at $w_{\rm th}$.

We computed the value of $w_{\rm th}$ for each of the typical target redshift ranges of the color techniques, i.e., $\rm LBG_{V09}$, BzK, VJL and iHM.The best $w_{\rm th}$ for each redshift range and each selection criterion are listed in Table 5.2 and indicated with a red line in Figure 5.6. In order to understand whether the efficiency of a given selection is depending on the type of sources, we have also used the $\rm LBG_{V09}$ and the BzK-like selections on non-X-ray sources and similarly computed w, C and P_s.

First of all, we found that the optimal value of $w_{\rm th}$ is frequently within the range of $0.3 < w < 0.5$ and this is valid for normal and X-ray detected galaxies. In addition, we note that in each redshift range, at the value of $w_{\rm th}$ that maximizes $(C + P_s)$, C and P_s are generally higher for X-ray sources than for non-X-ray normal galaxies. This implies that X-ray sources have more confident values of photo-z in the target redshift range. This is because X-ray sources usually have strong emission line features which can help in identifying the photo-z correctly, specially in the presence of Intermediate and Narrow band photoemtry. At $z \geq 4.5$, we do not have Xray sources; for normal galaxies the drop of the mean $w_{\rm th}$ at that redshift, is due to the fact that most of the emission lines shift to redder bands which have broader filter bandwidth so that the line flux contribution to the continuum is diluted. This, added to the increased photometric errors for faint sources increases the uncertainty of photo-z and reduces the weight, and also because of the small sample size, thus the trend become unstable at $z > 4$.

Following Table 5.2, we have selected samples by $w_{\rm th}$ that maximize $C + P_s$ (so called $P(z)$-selected samples) in the redshift ranges: (a) $1.4 < z < 2.5$; (b) $2.5 < z < 3.5$; (c) $3.5 < z < 4.5$; (d) $3.1 < z < 4.4$; (e) $3.4 < z < 4.5$; as defined from the literature. In the following we will compare these samples with those defined by the color techniques.

Table 5.2: The values of $w_{\rm th}$, completeness (C) and purity (P_s) for which ($C+P_s$) is maximum for X-ray and non-X-ray detected samples in Area 1.

	Non-X-ray				X-ray			
	$w_{\rm th}$	C	P_s	$(C+P_s)$	$w_{\rm th}$	C	P_s	$(C+P_s)$
$0.0 < z < 0.7$	0.50	0.77	0.91	1.68	0.50	0.98	1.00	1.98
$0.7 < z < 1.4$	0.40	0.76	0.78	1.54	0.50	0.95	0.98	1.93
$1.4 < z < 2.5$[a]	0.30	0.93	0.68	1.62	0.45	0.95	0.89	1.84
$2.5 < z < 3.5$[b]	0.45	0.64	0.80	1.44	0.25	0.93	0.65	1.58
$3.5 < z < 4.5$[c]	0.50	0.52	0.81	1.33	0.40	0.78	0.76	1.54
$3.1 < z < 4.4$[d]	0.45	0.66	0.83	1.49	0.45	0.85	0.91	1.76
$3.4 < z < 4.5$[e]	0.45	0.62	0.80	1.42	0.50	0.78	0.90	1.69

[a] Typical redshift range of BzK selection (Daddi et al., 2004).

[b] Typical redshift range of LBG selection for U-band dropout (Steidel et al., 1996); it is also the range of VJL selection from Guo et al. (2013)

[c] Redshift range of iHM selection from Guo et al. (2013)

[d] Redshift range of LBG selection for B-band dropouts from Vanzella et al. (2009)

[e] Redshift range of LBG selection for B-band dropouts from Papovich et al. (2004)

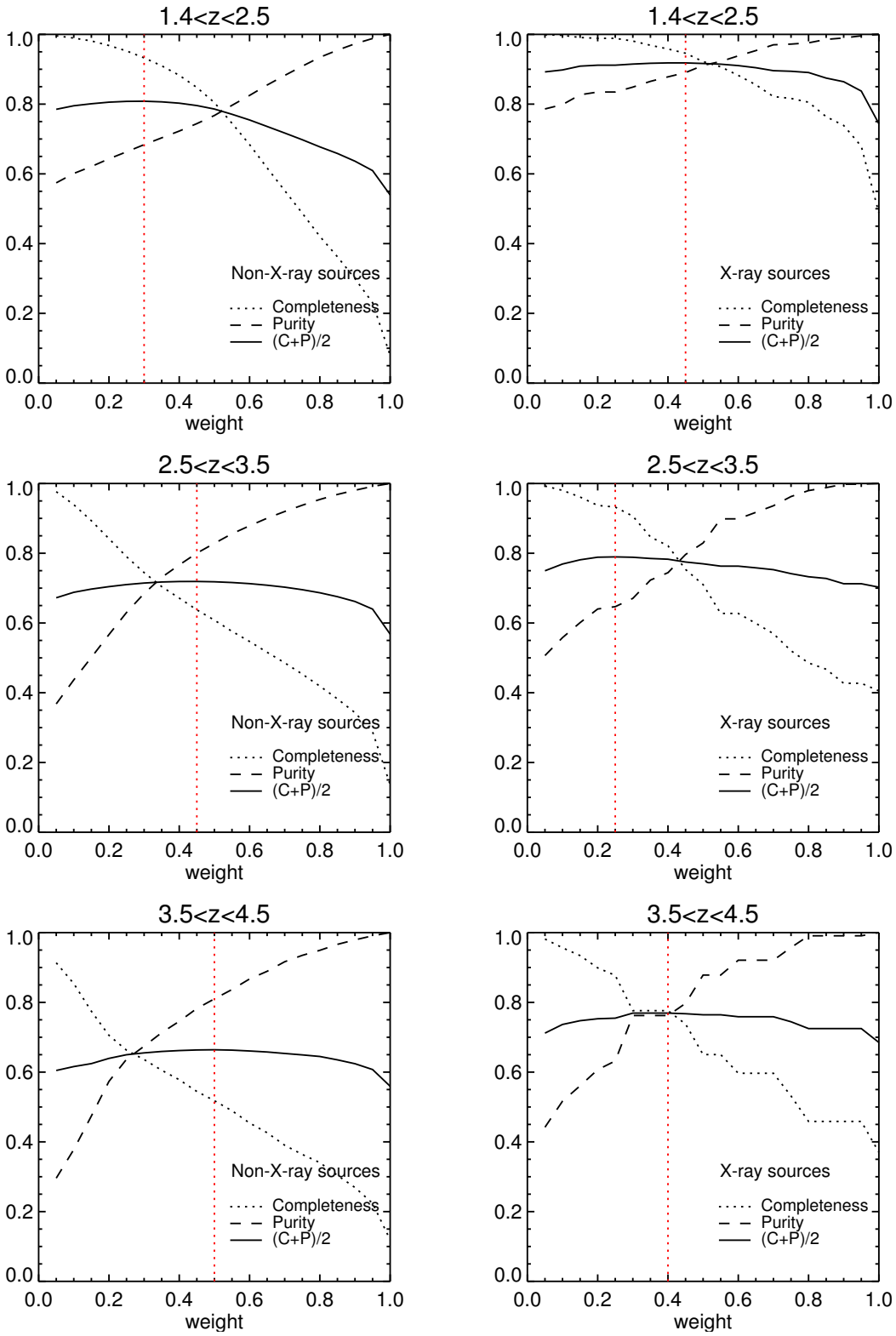


Figure 5.6: $(C + P_s)$ as a function of $w_{\rm th}$ at given redshift ranges of $1.4 < z < 2.5$, $2.5 < z < 3.5$, and $3.5 < z < 4.5$ in GOODS-S (Area 1) for non-X-ray (left panel) and X-ray (right panel) sources. The red dotted line represents the $w_{\rm th}$ selected by the maximum of the $(C + P_s)$.

5.4 Comparison between techniques for selecting high-z sources

To fully understand the color methods in selecting high-z sample, we estimate their completeness (C) and purity (P_s), using the value of ($C + P_s$) to compare their efficiencies with those measured for the $P(z)$ and the *zbest*-selected samples. In addition, we refine the redshift range in which each method maximizes its efficiency, comparing our optimal redshift ranges with those defined from the literature.

5.4.1 $\mathrm{LBG_{V09}}$ vs. $BzK/VJL/iHM$

In Section 5.1, we have defined color-selected samples by the $\mathrm{LBG_{V09}}$, BzK, VJL, and iHM methods. Here we make the comparison between them at their corresponding target redshift ranges. For the $\mathrm{LBG_{V09}}$ selection, three redshift ranges: $2.5 < z < 3.5$, $3.1 < z < 4.4$, and $3.4 < z < 4.5$, are defined in Steidel et al. (1996, 2003), Vanzella et al. (2009), and Papovich et al. (2004), respectively. Moreover, the range $1.4 < z < 2.5$, $2.5 < z < 3.5$, and $3.5 < z < 4.5$ are defined in Daddi et al. (2004) and Guo et al. (2013) for BzK, VJL, and iHM methods, respectively. We compare the ($C + P_s$) between the $\mathrm{LBG_{V09}}$, BzK, VJL and iHM methods in Table 5.3, which lists all the detailed values of samples in the five redshift ranges.

Among the three methods that used to select sources in the redshift range $2.5 < z < 4.5$ (i.e., $\mathrm{LBG_{V09}}$, VJL and iHM), the VJL technique has the highest efficiency ($C + P_s = 1.12$), with $\sim$ 30% purity. Similar purity is reached by the $\mathrm{LBG_{V09}}$ technique but its completeness is lower. In addition, although iHM has higher completeness than $\mathrm{LBG_{V09}}$, it has an extremely low purity of 9% which means that the sample is highly contaminated by low-redshift interlopers.

The completeness and purity of our $\mathrm{LBG_{V09}}$ samples are not as high as the samples defined in Steidel et al. (2003) who have found that $\mathrm{LBG_{V09}}$ selection at $z \sim 3$ is extremely clean for galaxies and only contaminated by very few low-redshift interlopers ($\sim 4\%$). In order to see whether this low efficiency is due to the fact that we are selecting AGNs but not galaxies, we have computed C and P_s also for non active galaxies confirming the same behaviors as for AGNs. It is interesting to note that completeness and purity are not the same among the LBG samples. For example, in the redshift range of $2.5 - 3.5$ the method is the one with the lower ($C + P_s$), while in the range $\sim 3 - 4.5$ (Vanzella et al., 2009; Papovich et al., 2004), have a similar efficiency (twice as good as Steidel et al. 2003) but with varying C and P_s among them.

For X-ray selected AGNs, Nandra et al. (2005), Aird et al. (2008) and Aird et al. (2010) also applied the LBG technique to constrain the X-ray luminosity function. They found that the $\mathrm{LBG_{S96}}$ selection for the $z \sim 3$ AGNs suffers little from the low-redshift contamination, and the majority of this interlopers are Galactic stars. However in our $\mathrm{LBG_{V09}}$-selected AGNs, the completeness is $\sim 25\%$, and $\sim 70\%$ of the sources are outside the optimal redshift range. This can be probably explained by the fact that the original

LBG_{S96} selection was designed carefully with custom filters by Steidel et al. (1996) and Steidel et al. (2003), and the particular depths is set to reduce low-redshift interlopers. In our work, instead we used HST/ACS and WFC3 filters, and this also can explain the low efficiency that we obtained. This result happens not only in the LBG selection, but also in other color methods. Table 5.4 compares the filter information in the original works with the modified versions that we used, as found in literature. In Section 5.4.4 we will use the maximized $(C + P_s)$ to define for each new filter set a new redshift range within which the various methods are efficient.

Table 5.3: Completeness (C) and purity (P_s) for AGN and galaxy samples selected by $P(z)$, *zbest* and LBG_{V09} $/BzK/VJL/iHM$ techniques in Area 1

			Galaxies				AGNs			
	Redshift range	Technique	N	C	P_s	$C+P_s$	N	C	P_s	$C+P_s$
LBG	$2.5 < z < 3.5$	LBG_{S96}	1868	0.13	0.31	0.44	31	0.23	0.33	**0.56**
		zbest	4195	0.67	0.73	1.39	44	0.76	0.76	1.52
		P(z)	3358	0.61	0.83	1.44	42	0.75	0.80	1.55
	$3.1 < z < 4.4$	LBG_{V09}	1868	0.37	0.35	0.72	31	0.54	0.37	**0.91**
		zbest	1421	0.66	0.80	1.45	21	0.87	0.88	1.75
		P(z)	1292	0.64	0.85	1.49	20	0.85	0.91	1.76
	$3.4 < z < 4.5$	LBG_{P04}	1868	0.49	0.23	0.72	31	0.60	0.29	**0.89**
		zbest	734	0.62	0.76	1.37	14	0.81	0.87	1.69
		P(z)	655	0.60	0.82	1.42	14	0.81	0.87	1.69
BzK	$1.4 < z < 2.5$	BzK	8838	0.46	0.69	1.15	177	0.80	0.67	**1.47**
		zbest	13674	0.80	0.78	1.58	154	0.94	0.91	1.85
		P(z)	14083	0.83	0.78	1.61	156	0.95	0.91	1.86
VJL	$2.5 < z < 3.5$	VJL	9940	0.72	0.33	1.04	114	0.81	0.31	**1.12**
		zbest	4195	0.67	0.73	1.39	44	0.76	0.76	1.52
		P(z)	3358	0.61	0.83	1.44	38	0.71	0.83	1.54
iHM	$3.5 < z < 4.5$	iHM	5749	0.70	0.08	0.78	88	0.72	0.09	**0.81**
		zbest	507	0.55	0.71	1.27	9	0.69	0.83	1.52
		P(z)	419	0.52	0.81	1.33	8	0.65	0.88	1.53

1. Note that the Galaxy and AGN here refer to the non-X-ray and X-ray detected samples respectively.
2. The efficiency, i.e., $C + P_s$, of each color method is presented in bold.
3. N: number of sources in the sample.

Table 5.4: Comparisons between the original works and our works on the filters, depth and regions used for color techniques.

	Original works		Our works	
	Filter information	Target redshift range	Filter information	Optimal redshift range by maximized $(C + P_s)$
LBG	Steidel et al. (2003) used U_nGR system from five ground-based telescopes in 17 fields, requiring magnitude limit: $19 < R < 25.5$	Galaxy: $z = 2.96 \pm 0.29$	B: F435W; V: F606W; z: F850LP	Galaxy: $3.3 < z < 4.2$ AGN: $3.3 < z < 3.8$
BzK	Daddi et al. (2004) used the data from K20 survey in GOODS-S, requiring magnitude limit: $K < 20$ B: VLT Bessel B-band + ACS F435W z: F850W K: VLT/ISSAC	Galaxy: $1.4 < z < 2.5$	B: F435W; z: F850LP; K: VLT/ISAAC	Galaxy: $1.0 < z < 3.4$ AGN:$1.2 < z < 3.2$
VJL	Guo et al. (2013) V: F606W; J: F125W; L: IRAC 3.6 μm	Galaxy: $2.5 < z < 3.5$	V: F606W J: F125W; L: IRAC 3.6 μm	Galaxy: $2.0 < z < 4.2$ AGN: $2.0 < z < 4.2$
iHM	Guo et al. (2013) i: F775W; H: F160W; M: IRAC 4.5μm	Galaxy: $3.5 < z < 4.5$	i: F775W; H: F160W; M: IRAC 4.5μm	Galaxy: $3.2 < z < 4.2$ AGN: $3.3 < z < 3.6$

B, V, z, V, J, i, H bands used in our work are from the CANDELS survey by HST/ACS +WFC3 with depth ~ 29 mag.
K band used in our work has depth ~ 25 mag
L and M bands are from the SEDS survey by *Spitzer* with depth ~ 26.5 mag
In our works, all the color selections are done in GOODS-S region.

5.4.2 $P(z)$ vs. $\mathrm{LBG_{V09}}$

We compared the $\mathrm{LBG_{V09}}$-selected AGNs (31 sources, see Section 5.1.1) with $P(z)$-selected sample (20 sources) in the redshift range $3.1 < z < 4.4$ (i.e., 6th row of Table 5.2 Vanzella et al., 2009). There are only 11 sources in common between them. Of the 9 sources that are picked up by $P(z)$ but not by LBG, all have zphot > 3 and 4 of them are also spectroscopically confirmed. The sources are also plotted in Figure 5.1 (blue circles). Most of them could be recovered by slightly modifying the $\mathrm{LBG_{V09}}$ wedge boundaries. This suggests that, perhaps, there is some issue with the translation of the original LBG method Steidel et al. (1996), which was tuned for a specific set of filters, to the new redshift range and new photometric bands, as adoped by Vanzella et al. (2009). It is important to keep in mind, however, that a larger wedge would increase also the fraction of low-z interlopers.

Table 5.5: Properties of $P(z)$-selected sample that are not included in the $\mathrm{LBG_{V09}}$ sample in Area 1.

$\mathrm{ID_{H14}}$ (1)	zbest (2)	$w(3.1 < z < 4.4)$ (3)	$w(3.3 < z < 3.8)$ (4)	mod (5)
20936	3.256*	1.0	0.0	16
23382	4.379*	1.0	0.0	42
3973	3.876	0.97	0.56	2
13327	3.2405	0.86	0.60	44
22426	3.2471	0.77	0.03	16
16822	3.2419	0.63	0.34	9
4624	3.173	0.5	0.25	42
3360	3.471*	1.0	1.0	106
4936	3.61*	1.0	1.0	128

NOTE: The column description is the same as Table 5.1.

5.4.3 $P(z)$ vs. $BzK/VJL/iHM$

As for $\mathrm{LBG_{V09}}$ sample in Section 5.1.1, we have compared the overlap between the $P(z)$-selected sample with the samples selected via BzK, VJL, and iHM color methods and summarized the results in Table 5.3. Due to the large number of sources involved in the analysis, we do not present detailed tables as we have done for $\mathrm{LBG_{V09}}$ in Table 5.5. We found that the BzK, VJL and iHM selected samples have very high completenesses and largely overlap with $P(z)$-selected sample.

In particular the VJL/iHM-selected samples are more complete than $P(z)$. However, the purity for all three methods drop steeply from the BzK to iHM samples. These methods are strongly biased toward low-redshift interlopers, and the majority of the sources in the VJL/iHM-selected samples are outside the targeted redshift range (i.e., $2.5 < z \leq$

3.5, $3.5 < z \leq 4.5$ respectively, as defined in Guo et al. 2013). In particular for the iHM method, more than 90% of the sources have redshift outside the original range of interest. This suggest that at least the iHM as defined in Guo et al. (2013) can not be used to select reliable high-z samples of AGN.

5.4.4 Optimizing the efficiency of color selection techniques defining high-z sources

From the discussions above, it can be confirmed that the efficiency of each color selection depends on the redshift ranges under consideration. In this section, we have redefined, for each method, the range of redshift in which they perform best, in terms of maximum value $(C + P_s)$. In Table 5.6 we listed the results we obtained for both AGNs and galaxies.

To better understand the comparisons between the newly-defined redshift range with previous works, we visualized the results summarized in Table 5.3 and Table 5.6, plotting the efficiency for each color-based method as a function of the target redshift range in Figure 5.7. For the $\mathrm{LBG_{V09}}$-selected galaxies, the optimal redshift of $3.3 < z < 4.2$ is similar to the ranges defined by Papovich et al. (2004) and Vanzella et al. (2009), while selecting the high-z AGNs via the $\mathrm{LBG_{V09}}$ technique is improved only if performed within the much smaller range of $3.3 < z < 3.8$. On the contrary, selecting galaxies and AGNs using the BzK and VJL is more efficient if extended to a larger ranges than those defined originally by Daddi et al. (2004) and Guo et al. (2013).

An important result of the analysis presented in Table 5.3 and Table 5.6, is that the $P(z)$ is the method produces the most complete and pure samples at *any* redshift.

Table 5.6: Redshift range optimized by maximizing $(C + P_s)$ for AGN and galaxy samples selected by $P(z)$, *zbest* or $\mathrm{LBG_{V09}}/BzK/VJL/iHM$ techniques in Area 1

	Redshift range	Technique	N	C	P_s	$C + P_s$
LBG	Galaxy	$\mathrm{LBG_{V09}}$	1868	0.48	0.27	0.76
	$3.3 < z < 4.2$	*zbest*	892	0.65	0.77	1.42
		P(z)	773	0.62	0.85	1.47
	AGN	$\mathrm{LBG_{V09}}$	31	0.67	0.30	0.98
	$3.3 < z < 3.8$	*zbest*	14	0.83	0.83	1.66
		P(z)	14	0.85	0.85	1.71
BzK	Galaxy	BzK	8838	0.36	0.93	1.29
	$1.0 < z < 3.4$	*zbest*	22339	0.92	0.93	1.84
		P(z)	23453	0.95	0.91	1.87
	AGN	BzK	177	0.75	0.88	1.63
	$1.2 < z < 3.2$	*zbest*	206	0.97	0.98	1.95
		P(z)	209	0.98	0.97	1.95
VJL	Galaxy	VJL	9940	0.61	0.60	1.21
	$2.0 < z < 4.2$	*zbest*	8887	0.73	0.80	1.53
		P(z)	8590	0.73	0.82	1.55
	AGN	VJL	114	0.73	0.67	1.41
	$2.0 < z < 4.2$	*zbest*	109	0.93	0.89	1.82
		P(z)	111	0.94	0.89	1.83
iHM	Galaxy	iHM	5749	0.70	0.16	0.86
	$3.2 < z < 4.2$	*zbest*	1101	0.65	0.78	1.43
		P(z)	970	0.62	0.85	1.47
	AGN	iHM	88	0.97	0.09	1.06
	$3.3 < z < 3.6$	*zbest*	8	0.71	0.72	1.43
		P(z)	6	0.63	0.86	1.49

N: number of sources in the sample.
C: completeness.
P_s: purity.

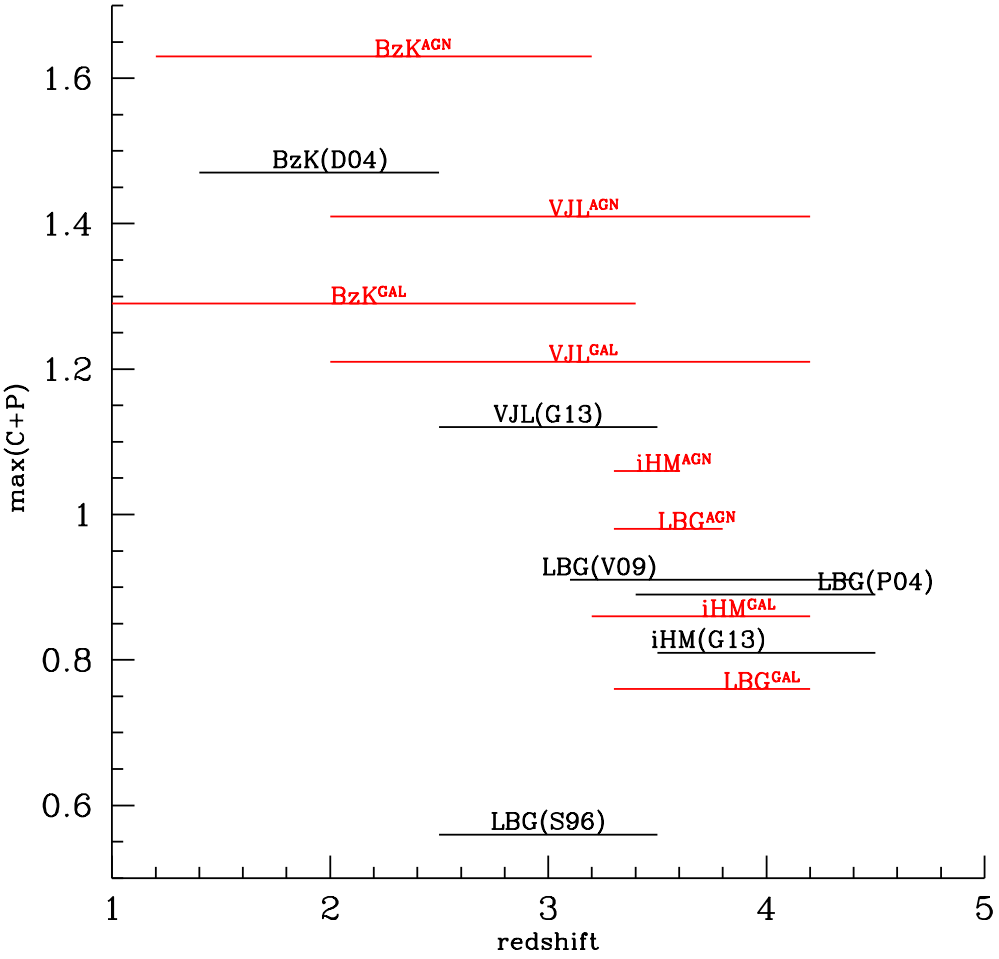


Figure 5.7: Selection efficiencies of color methods as a function of type of sources and redshift ranges. Black lines are the ranges defined by the original works, and red lines are the ranges refined from our work. V09:Vanzella et al. (2009); P04: Papovich et al. (2004); D04: Daddi et al. (2004); G13: Guo et al. (2013); S96: Steidel et al. (1996).

5.5 High-redshift ($z > 3$) candidates selected by $P(z)$

5.5.1 Sample selection

Starting from the 1259 X-ray sources from the 4Ms-CDFS and 250ks-ECDFS catalogs, we selected the sample at $z > 3$ using both spec-z and photo-z. There are 20 sources with reliable spec-z ≥ 3, 53 sources have photo-z ≥ 3, and 79 sources with *zbest* < 3 but $z_{\rm p} + 3\sigma \geq 3$. Table 5.7 summarizes the number of high-z candidates with spec-z and photo-z in our three areas.

Table 5.7: Number of high-redshift ($z \geq 3$) X-ray sources

Field	$z_{\rm s} \geq 3$	$z_{\rm p} \geq 3$	*zbest* < 3 and $z_{\rm p} + 3\sigma \geq 3$	Total	number of sources with $w \geq 0.45$
Area 1	10	16	28	54	23
Area 2	2	14	19	35	17
Area 3	8	23	32	63	25
Area 1+2+3	20	53	79	152	65

However, as discussed in the previous section, a more reliable high-z sample can be based on the $P(z)$ technique and the probability of the sources being at high redshift as defined by the weight w. For this reason, we computed w for the 152 sources by computing the fraction of area beneath the $P(z)$ curve at $z > 3$ (i.e., blue shaded area in the right plots of Figure 5.8) via the equation below:

$$w(z \geq 3) = \frac{\int_3^7 P(z)dz}{\int_0^7 P(z)dz} \tag{5.4}$$

The value of $w(z \geq 3)$ represents how confident is a source having a photo-z ≥ 3 ($w = 1$ for sources with spec-z). Figure 5.8 shows three examples of high-redshift candidates. In general, with sufficient data points, we can constrain better the redshift solution (i.e., peaked $P(z)$, the first example), whereas with fewer data points, the $P(z)$ can be very broad, implying a less constrained photo-z (the second example). High-redshift candidates could also have more than one peak value (as presented in the third example, showing two peaks at $z \sim 0.7$ and $z \sim 3$).

Similar to what has been done in the previous sections, we can compute the value of w that maximizes $(C + P_s)$ for sources at $z > 3$. The values is $w(z \geq 3) = 0.45$. 55 out of 73 ($\sim 75\%$) sources have *zbest* > 3 with $w(z \geq 3)$ greater than 0.45 (see detail in Table 5.8). For the other 79 sources with *zbest* < 3 and $(z_{\rm p} + 3\sigma) \geq 3$, only 10 ($\sim 13\%$) sources have $w(z \geq 3) > 0.45$ (see detail in Table 5.9). In both tables, we sorted the sources by the values of $w(z \geq 3)$. Now that we have a well defined sample we can compared with a similar sample defined in the same area (CANDLES/GOODS-S) by use of *zbest* (Fiore et al., 2012).

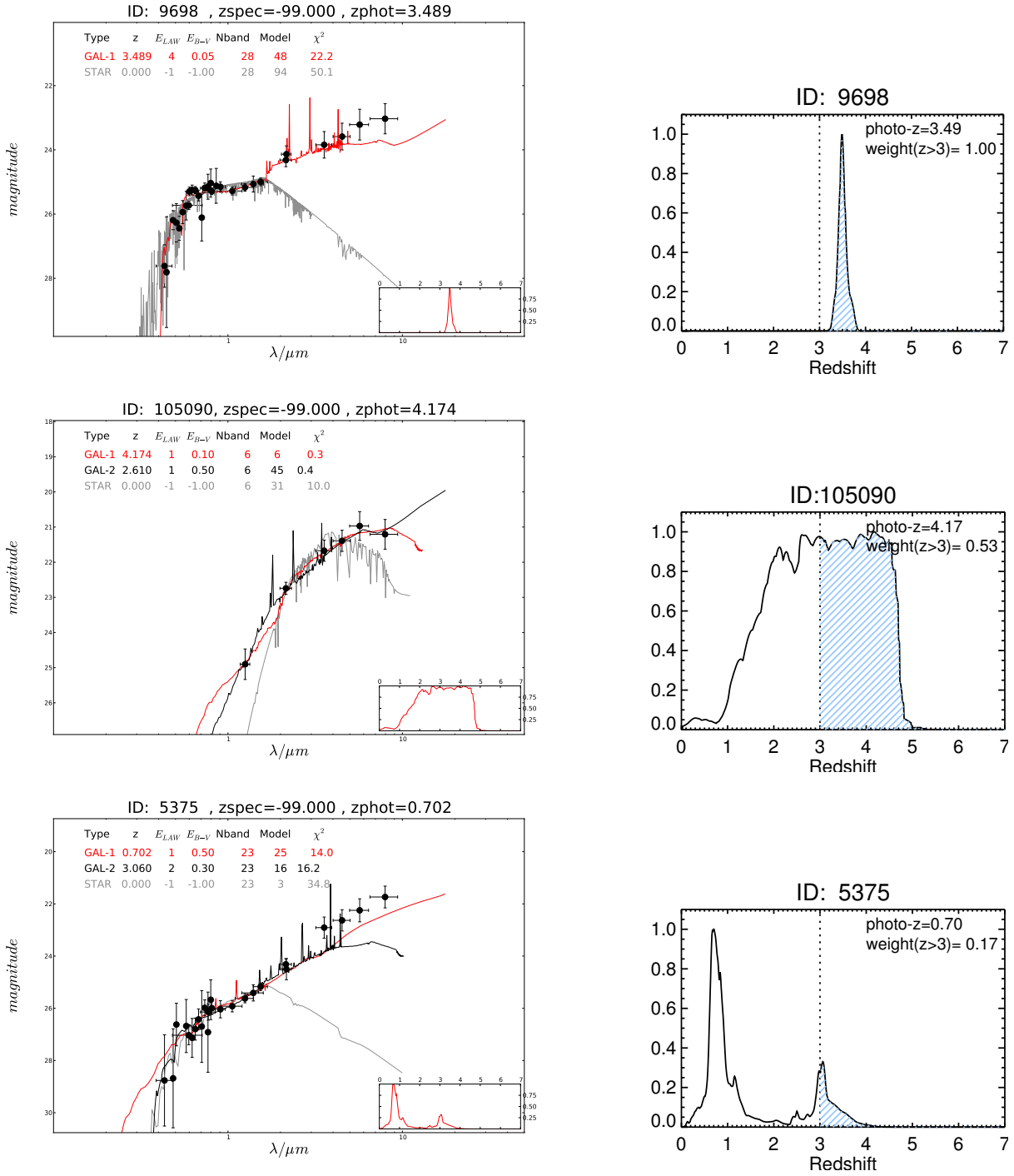


Figure 5.8: Left panel: Examples of SEDs of high-z candidates. Right panel: $P(z)$ curves of high-z candidates; blue shades are the area at $z > 3$.

Table 5.8: High-redshift source list with $zbest > 3$

ID_{H14}	ID_{R13}	ID_{X11}	ID_{L05}	ID_{V06}	Area	$zbest$	z_s	Q_{zs}	z_p	$3\sigma_{z_p}$	$w(z \geq 3)$
105117	138	674	-99	-99	2	3.082	3.082	0	3.143	(0.58-5.03)	1.0
91706	489	361	-99	-99	2	3.544	-99.0	-99	3.544	(3.50-3.58)	1.0
91352	-99	-99	108	203	3	3.0	3.0	0	2.953	(2.94-2.98)	1.0
87475	-99	-99	45	215	3	3.408	3.408	0	3.428	(3.41-3.44)	1.0
87044	-99	-99	660	24	3	3.5	3.5	0	3.464	(3.44-3.49)	1.0
78608	-99	-99	51	213	3	4.48	4.48	0	4.5	(4.47-4.53)	1.0
66174	406	10	-99	-99	2	3.153	3.153	1	2.532	(2.48-2.63)	1.0
58153	-99	-99	603	429	3	3.63	3.63	0	2.635	(2.61-2.67)	1.0
48250	-99	-99	285	272	3	3.381	3.381	0	0.393	(0.38-0.40)	1.0
42881	-99	-99	408	481	3	3.16	3.16	0	1.384	(1.35-1.42)	1.0
41671	-99	-99	638	420	3	3.092	3.092	0	3.074	(3.05-3.10)	1.0
39682	-99	-99	630	493	3	4.803	-99.0	-99	4.803	(4.79-4.82)	1.0
23382	334	534	-99	-99	1	4.379	4.379	0	2.425	(1.96-2.56)	1.0
20936	257	386	-99	-99	1	3.256	3.256	0	2.985	(2.40-3.07)	1.0
14587	110	573	-99	-99	1	3.193	3.193	1	3.222	(3.18-3.28)	1.0
10578	87	546	-99	-99	1	3.064	3.064	0	2.835	(2.66-3.09)	1.0
8833	311	132	-99	-99	1	3.515	-99.0	-99	3.515	(3.31-3.63)	1.0
6294	47	412	-99	-99	1	3.7	3.7	0	3.695	(3.65-3.74)	1.0
5479	39	262	-99	-99	1	3.66	3.66	0	3.646	(3.60-3.70)	1.0
4936	33	563	-99	-99	1	3.61	3.61	0	3.562	(3.51-3.58)	1.0
3372	18	645	-99	-99	1	3.47	3.47	2	3.521	(0.34-3.58)	1.0
3360	19	588	-99	-99	1	3.471	3.471	0	3.43	(3.41-3.46)	1.0
273	-99	403	-99	-99	1	4.76	4.76	1	4.695	(4.53-4.86)	1.0
48101	-99	-99	485	457	3	3.151	-99.0	-99	3.151	(3.11-3.19)	0.999
9698	408	458	-99	-99	1	3.489	-99.0	-99	3.489	(3.25-3.80)	0.999
6905	229	235	-99	-99	1	3.639	-99.0	-99	3.639	(3.45-3.80)	0.999
19505	323	100	-99	-99	1	3.417	-99.0	-99	3.417	(3.24-3.95)	0.988
3973	23	460	-99	-99	1	3.876	-99.0	-99	3.876	(3.05-4.28)	0.98
62806	222	700	-99	-99	2	3.631	-99.0	-99	3.631	(3.45-3.79)	0.969
13327	107	430	-99	-99	1	3.24	-99.0	-99	3.24	(2.38-4.02)	0.92
12309	-99	150	-99	-99	1	3.438	-99.0	-99	3.438	(2.77-3.69)	0.905
24636	486	602	-99	-99	1	3.7	-99.0	-99	3.7	(0.23-3.98)	0.884
67889	-99	-99	-99	542	3	3.997	-99.0	-99	3.997	(0.13-5.26)	0.852
59283	-99	-99	706	577	3	4.627	-99.0	-99	4.627	(0.78-4.66)	0.797
29323	231	156	-99	-99	1	4.651	-99.0	-99	4.651	(0.19-7.00)	0.795
22426	-99	554	-99	-99	1	3.247	-99.0	-99	3.247	(0.15-3.33)	0.785
105145	192	139	-99	-99	2	5.226	-99.0	-99	5.226	(0.45-5.58)	0.725
5367	-99	374	-99	-99	1	3.519	-99.0	-99	3.519	(0.22-3.76)	0.718
16822	130	371	-99	-99	1	3.242	-99.0	-99	3.242	(0.37-4.53)	0.711
105045	-99	-99	569	436	3	5.11	-99.0	-99	5.11	(0.01-7.00)	0.707
33160	444	85	-99	-99	1	3.365	-99.0	-99	3.365	(0.04-7.00)	0.688
105012	-99	-99	323	641	3	4.234	-99.0	-99	4.234	(0.00-5.68)	0.677
102535	-99	-99	330	247	3	3.104	-99.0	-99	3.104	(2.24-3.17)	0.636
105010	-99	-99	545	503	3	4.125	-99.0	-99	4.125	(0.00-7.00)	0.632
105079	353	69	-99	-99	2	3.645	-99.0	-99	3.645	(0.53-7.00)	0.621
4624	32	456	-99	-99	1	3.173	-99.0	-99	3.173	(1.87-4.57)	0.618
105051	285	230	-99	-99	2	3.985	-99.0	-99	3.985	(0.87-6.40)	0.617
105124	-99	-99	551	49	3	5.034	-99.0	-99	5.034	(0.00-7.00)	0.594
105094	-99	730	-99	-99	2	6.648	-99.0	-99	6.648	(0.00-7.00)	0.586

Table 5.8: (Continued)

ID_{H14}	ID_{R13}	ID_{X11}	ID_{L05}	ID_{V06}	Area	*zbest*	z_s	Q_{zs}	z_p	$3\sigma_{z_p}$	$w(z \geq 3)$
105095	455	703	-99	-99	2	4.685	-99.0	-99	4.685	(0.00-7.00)	0.579
5064	402	258	-99	-99	1	3.05	-99.0	-99	3.05	(2.02-4.00)	0.567
105128	-99	-99	33	218	3	4.783	-99.0	-99	4.783	(0.00-6.80)	0.556
92315	386	253	-99	-99	2	3.315	-99.0	-99	3.315	(1.19-3.41)	0.54
105090	-99	44	-99	-99	2	4.174	-99.0	-99	4.174	(0.15-4.96)	0.528
105028	-99	-99	84	-99	3	3.501	-99.0	-99	3.501	(0.19-5.01)	0.525
105064	-99	-99	650	418	3	4.955	-99.0	-99	4.955	(0.00-5.82)	0.523
105070	-99	-99	10	394	3	4.325	-99.0	-99	4.325	(0.00-5.66)	0.493
105025	-99	-99	215	293	3	3.872	-99.0	-99	3.872	(0.87-4.74)	0.482
105133	-99	-99	30	-99	3	3.114	-99.0	-99	3.114	(0.99-4.53)	0.481
105022	-99	-99	63	341	3	3.545	-99.0	-99	3.545	(0.00-5.83)	0.48
105050	-99	-99	131	318	3	3.116	-99.0	-99	3.116	(0.71-4.93)	0.479
105105	361	715	-99	-99	2	3.234	-99.0	-99	3.234	(0.51-4.69)	0.467
55958	-99	-99	86	335	3	3.124	-99.0	-99	3.124	(1.85-3.17)	0.455
105093	470	64	-99	-99	2	3.15	-99.0	-99	3.15	(0.97-3.87)	0.424
105089	85	707	-99	-99	2	3.639	-99.0	-99	3.639	(0.75-4.12)	0.417
105016	-99	-99	481	459	3	3.197	-99.0	-99	3.197	(0.75-4.08)	0.411
105134	-99	-99	696	13	3	3.692	-99.0	-99	3.692	(0.58-4.22)	0.398
105069	217	129	-99	-99	2	3.446	-99.0	-99	3.446	(0.38-3.87)	0.35
105035	-99	-99	709	615	3	3.114	-99.0	-99	3.114	(0.79-3.87)	0.335
105036	-99	-99	219	292	3	3.163	-99.0	-99	3.163	(0.00-4.43)	0.332
105047	282	524	-99	-99	2	3.096	-99.0	-99	3.096	(0.76-3.70)	0.315
26315	551	272	-99	-99	1	3.019	-99.0	-99	3.019	(0.28-4.09)	0.257
81619	-99	-99	12	244	3	5.35	-99.0	-99	5.35	(0.70-5.36)	0.193

The first five columns are the IDs from this work, R13, X11, L05, and V06, respectively.
Area: area number, see Figure 2.1.
zbest: spec-z, otherwise photo-z, i.e., the peak value of $P(z)$.
z_s: spec-z
Q_{zs}: quality of the spec-z (0=high, 1=good, 2=intermediate, 3=poor, and -99=no spec-z).
z_p: photo-z, i.e., the peak value of $P(z)$. .
$3\sigma_{z_p}$: 3σ range of photo-z.
$w(z \geq 3)$: weight of photo-z at $z \geq 3$. We set this value = 1 for the spec-z.

Table 5.9: High-redshift source list with $zbest < 3$ and $(z_{\rm p} + 3\sigma) \geq 3$

ID_{H14}	ID_{R13}	ID_{X11}	ID_{L05}	ID_{V06}	Area	$z_{\rm best}$	$z_{\rm s}$	$Q_{\rm zs}$	$z_{\rm p}$	$3\sigma_{z_{\rm p}}$	$w(z \geq 3)$
105100	247	713	-99	-99	2	2.167	-99.0	-99	2.167	(0.00-7.00)	0.62
105088	73	6	-99	-99	2	2.005	-99.0	-99	2.005	(0.00-7.00)	0.619
105041	-99	610	-99	-99	2	2.564	-99.0	-99	2.564	(0.00-7.00)	0.582
105021	-99	-99	6	359	3	1.673	-99.0	-99	1.673	(0.00-7.00)	0.554
91565	-99	-99	442	-99	3	2.982	-99.0	-99	2.982	(2.91-3.14)	0.52
7061	542	630	-99	-99	1	2.984	-99.0	-99	2.984	(0.97-5.17)	0.518
105046	5	197	-99	-99	2	2.881	-99.0	-99	2.881	(0.11-5.27)	0.502
105130	-99	-99	512	61	3	2.511	-99.0	-99	2.511	(1.03-4.24)	0.461
105073	351	7	-99	-99	2	1.927	-99.0	-99	1.927	(0.19-4.71)	0.459
105087	70	689	-99	-99	2	1.267	-99.0	-99	1.267	(0.00-7.00)	0.459
105082	304	59	-99	-99	2	1.795	-99.0	-99	1.795	(0.00-5.65)	0.447
105101	117	105	-99	-99	2	1.26	-99.0	-99	1.26	(0.00-7.00)	0.434
105032	-99	-99	633	-99	3	2.213	-99.0	-99	2.213	(0.67-4.69)	0.415
105006	-99	-99	77	337	3	2.58	-99.0	-99	2.58	(0.36-4.43)	0.386
104981	-99	-99	38	608	3	2.943	-99.0	-99	2.943	(0.00-5.81)	0.383
105120	-99	-99	432	77	3	2.262	-99.0	-99	2.262	(0.00-4.61)	0.382
105068	22	27	-99	-99	2	2.727	-99.0	-99	2.727	(0.00-4.79)	0.375
4356	28	485	-99	-99	1	2.475	-99.0	-99	2.475	(0.18-4.93)	0.355
105149	365	3	-99	-99	2	0.833	-99.0	-99	0.833	(0.00-4.38)	0.349
105135	-99	-99	724	588	3	1.774	-99.0	-99	1.774	(0.84-4.75)	0.325
104990	-99	-99	49	344	3	2.08	-99.0	-99	2.08	(0.00-5.52)	0.308
105114	-99	-99	270	237	3	1.502	-99.0	-99	1.502	(0.00-4.36)	0.306
77852	-99	-99	682	-99	3	0.605	-99.0	-99	0.605	(0.41-4.35)	0.303
105077	223	8	-99	-99	2	2.926	-99.0	-99	2.926	(0.00-3.97)	0.281
19713	522	392	-99	-99	1	2.402	-99.0	-99	2.402	(1.13-4.01)	0.276
105008	-99	-99	531	-99	3	2.922	-99.0	-99	2.922	(0.00-3.94)	0.239
4854	221	276	-99	-99	1	1.519	-99.0	-99	1.519	(1.02-4.18)	0.218
105048	-99	-99	2	361	3	1.73	-99.0	-99	1.73	(0.00-4.00)	0.199
5375	-99	331	-99	-99	1	0.702	-99.0	-99	0.702	(0.19-3.73)	0.169
105081	-99	-99	718	526	3	2.126	-99.0	-99	2.126	(0.00-3.79)	0.167
104983	-99	-99	574	643	3	1.559	-99.0	-99	1.559	(0.89-3.69)	0.153
105053	-99	-99	37	377	3	1.319	-99.0	-99	1.319	(0.25-3.56)	0.149
18969	-99	642	-99	-99	1	0.502	-99.0	-99	0.502	(0.20-3.78)	0.137
104997	-99	-99	654	502	3	1.849	-99.0	-99	1.849	(0.97-3.59)	0.129
5652	-99	285	-99	-99	1	1.719	-99.0	-99	1.719	(0.32-3.37)	0.109
19883	326	90	-99	-99	1	2.095	-99.0	-99	2.095	(0.35-3.66)	0.108
104995	-99	-99	-99	395	3	1.057	-99.0	-99	1.057	(0.45-3.88)	0.101
104996	-99	-99	384	650	3	1.057	-99.0	-99	1.057	(0.45-3.88)	0.101
4438	219	651	-99	-99	1	2.446	-99.0	-99	2.446	(1.13-3.41)	0.095
5827	43	528	-99	-99	1	2.973	-99.0	-99	2.973	(2.91-3.03)	0.092
53550	-99	-99	111	-99	3	1.275	-99.0	-99	1.275	(0.19-3.55)	0.084
48123	-99	-99	733	-99	3	0.401	-99.0	-99	0.401	(0.16-3.32)	0.08
26358	269	341	-99	-99	1	1.34	-99.0	-99	1.34	(1.12-4.22)	0.075
105002	-99	-99	299	382	3	0.995	-99.0	-99	0.995	(0.00-3.67)	0.065
89002	-99	-99	629	30	3	2.076	-99.0	-99	2.076	(0.73-3.22)	0.064
18694	534	177	-99	-99	1	0.608	-99.0	-99	0.608	(0.20-3.32)	0.063
105044	-99	-99	201	296	3	1.432	-99.0	-99	1.432	(0.52-3.26)	0.059
10294	80	466	-99	-99	1	2.775	-99.0	-99	2.775	(2.25-3.16)	0.056
6131	46	298	-99	-99	1	1.813	-99.0	-99	1.813	(1.19-3.26)	0.055

Table 5.9: (Continued)

ID_{H14}	ID_{R13}	ID_{X11}	ID_{L05}	ID_{V06}	Area	z_{best}	z_s	Q_{zs}	z_p	$3\sigma_{z_p}$	weight
48084	-99	-99	487	-99	3	1.873	-99.0	-99	1.873	(0.18-3.28)	0.054
13133	413	189	-99	-99	1	1.646	-99.0	-99	1.646	(1.00-3.46)	0.038
63777	227	60	-99	-99	2	2.509	-99.0	-99	2.509	(0.57-3.30)	0.037
57052	-99	-99	594	-99	3	1.392	-99.0	-99	1.392	(0.30-3.42)	0.036
2667	-99	496	-99	-99	1	2.496	-99.0	-99	2.496	(1.13-3.19)	0.035
11834	97	301	-99	-99	1	2.469	-99.0	-99	2.469	(1.11-3.22)	0.032
105054	-99	-99	676	412	3	2.713	-99.0	-99	2.713	(0.95-3.16)	0.032
104986	-99	-99	48	612	3	1.795	-99.0	-99	1.795	(0.00-3.49)	0.031
57711	-99	-99	711	405	3	2.4	-99.0	-99	2.4	(0.28-3.26)	0.03
105062	12	110	-99	-99	2	2.0	-99.0	-99	2.0	(0.18-3.13)	0.026
5867	44	354	-99	-99	1	2.682	-99.0	-99	2.682	(2.16-3.09)	0.025
20765	550	521	-99	-99	1	2.59	-99.0	-99	2.59	(0.53-3.15)	0.022
84794	197	600	-99	-99	2	1.766	-99.0	-99	1.766	(0.40-3.14)	0.021
26558	367	79	-99	-99	1	1.905	-99.0	-99	1.905	(1.16-3.13)	0.018
3943	524	348	-99	-99	1	2.608	-99.0	-99	2.608	(1.13-3.17)	0.017
135	281	469	-99	-99	1	2.643	-99.0	-99	2.643	(1.12-3.04)	0.016
10323	82	195	-99	-99	1	1.452	-99.0	-99	1.452	(0.25-3.08)	0.016
21227	-99	104	-99	-99	1	1.109	-99.0	-99	1.109	(0.21-3.18)	0.015
104991	-99	-99	409	582	3	2.541	-99.0	-99	2.541	(0.00-3.29)	0.015
9672	-99	328	-99	-99	1	2.425	-99.0	-99	2.425	(1.15-3.09)	0.014
61916	29	170	-99	-99	2	1.2	-99.0	-99	1.2	(0.39-3.07)	0.014
64006	377	92	-99	-99	2	2.164	-99.0	-99	2.164	(0.99-3.12)	0.013
3727	-99	303	-99	-99	1	2.06	-99.0	-99	2.06	(0.72-3.07)	0.011
105076	301	680	-99	-99	2	2.697	-99.0	-99	2.697	(0.15-3.24)	0.009
105112	-99	-99	421	629	3	1.522	-99.0	-99	1.522	(0.49-3.14)	0.008
24367	-99	475	-99	-99	1	2.134	-99.0	-99	2.134	(1.54-3.02)	0.007
2898	494	422	-99	-99	1	2.492	-99.0	-99	2.492	(1.17-3.02)	0.005
105096	-99	-99	4	245	3	1.443	-99.0	-99	1.443	(0.00-3.06)	0.004
51411	567	-99	-99	-99	2	2.736	-99.0	-99	2.736	(0.25-3.01)	0.003
83286	-99	103	-99	-99	2	2.163	-99.0	-99	2.163	(0.82-3.00)	0.002

Note: The column description is the same as in Table 5.8

5.5.2 Comparison with previous results

Fiore et al. (2012) discussed a sample of 54 X-ray sources at $z > 3$. The search criteria was not determined starting from the X-ray detection but rather by searching X-ray sources at the position of optical/NIR selected sources having $zbest > 3$ in the GOODS-ERS (Grazian et al., 2011; Santini et al., 2012) and the GOODS-MUSIC surveys (Grazian et al., 2006b; Santini et al., 2009). In this way they hoped to get fainter X-ray detections at high-z that were missed by the usual X-ray data analysis. They found that out of 54 sources, 29 were newly-found X-ray detections.

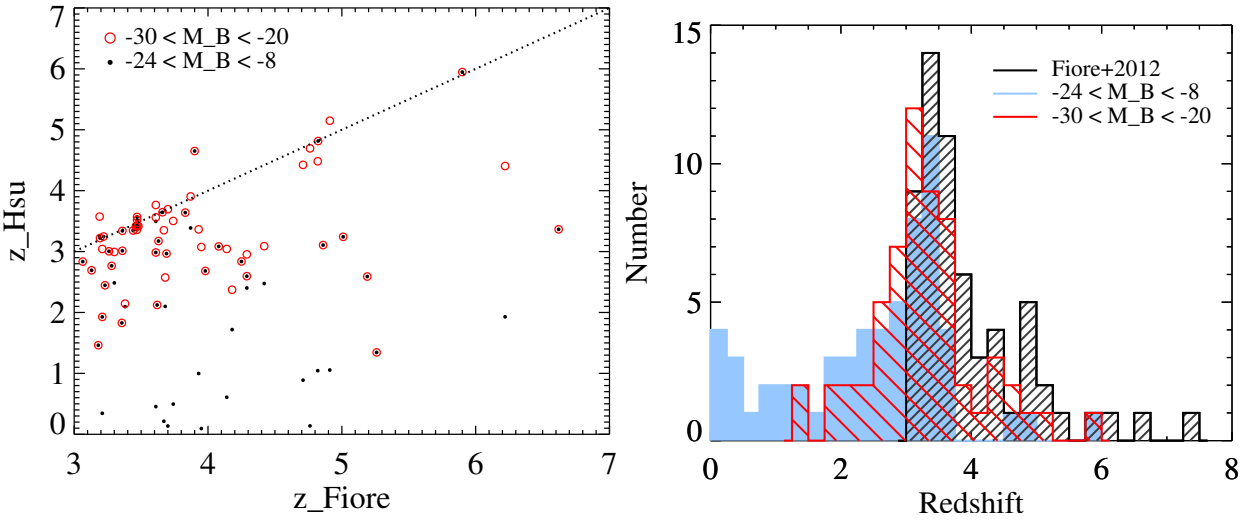
We compared their sample with our high-z sample determined in Section 5.5 by adopting $w(z \geq 3) > 0.45$ as weight threshold. We found that 29 X-ray sources ($\sim 48\%$) are in common and 19 of them have comparable redshifts (redshift differences are less than 0.1). Fourteen out of 19 are also spectroscopically confirmed. The discrepancies on the other sources are mainly due to two factors:

- First, the photo-z of the initial high-z optical sample used by the authors, is determined by using only normal galaxy templates that, at those redshifts rely on the Lyman break feature, independent of the type of source (i.e., galaxy or AGN dominated). Instead, we argue that depending on the template and priors adopted, the photo-z results can be very different, as already demonstrated also in Salvato et al. (2009, 2011).

  In our work, we adopted priors of $-30 < M_B < -20$ for point-like sources and $-24 < M_B < -8$ for extended (see detail in Section 3.2). For assessing how this can effect our comparison with Fiore et al. (2012), we re-computed the photo-z for their 54 $z > 3$ sources adopting both priors separately and see how the result changes (see the detailed result in Table 5.10 and Table 5.11). Figure 5.9a shows that with our templates, tuned for AGN, would provide in general redshifts that are lower than those assumed by Fiore et al. (2012).

- The second assumption is that the high-z candidate optical/NIR positions are unequivocally associated with the X-ray emission. We have already demonstrated that even with resolution of WFC3 images, we are not able to find a unique counterpart to the X-ray sources in all cases. The problem of getting the right X-ray-to-optical associations gets even more severe when using ground-based images for the source detection as in GOODS-MUSIC used by Fiore et al. (2012).

  In the following, we show two examples for clarifying the problem. In the first case (see Figure 5.10) we show two H-band detected sources ($\mathrm{ID_{H14}}$: 3372, 3395) that from the Bayesian analysis appear to be both highly likely (p >0.98) counterparts associated to the X-ray detection (X-645). However the two sources have completely different redshift (source 3372 has spec-z = 3.47, while source 3395 has a spec-z = 0.665).



(a) (b)

Figure 5.9: Left: Comparison between the photo-z assumed by Fiore et al. (2012) and those that we obtained by adopting templates of AGN, using different priors. For most of the sources our method estimated lower redshift values. The dotted line represents the 1-to-1 relation. Right: Same data as on the left panel, but this time summarized as a histogram.

In the second example (see panel (e), (g), (i) of Figure 5.11) we show the specific case of 7 X-ray detected sources that Fiore et al. (2012) assume to be associated to an optical counterpart at redshift z>5 (see Table 5.12). Our Bayesian approach suggest that the counterpart is likely a sources lying close to the X-ray position, but with a much lower redshift.

Finally in the same figure we show a further problem, namely the risk of using pure value of *zbest* in determining a high redshift sample of sources. For most of the seven cases, the $P(z)$ is not peaked at a single value but rather a broad range of possible redshifts. This strengthens our claim that a better sample can be defined by use of P(z) and w.

Lastly, compared to the high-z sample found by Fiore et al. (2012), we found 10 more X-ray sources at $z > 3$ in the CANDELS/GOODS-S region using our method. Most of them have $w(z \geq 3)$ greater than 0.45, and are likely to be at $z > 3$ (Table 5.13).

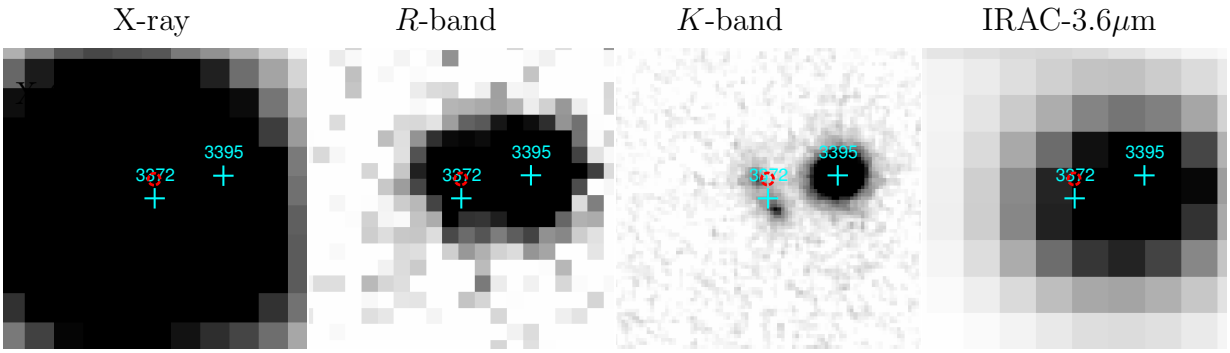
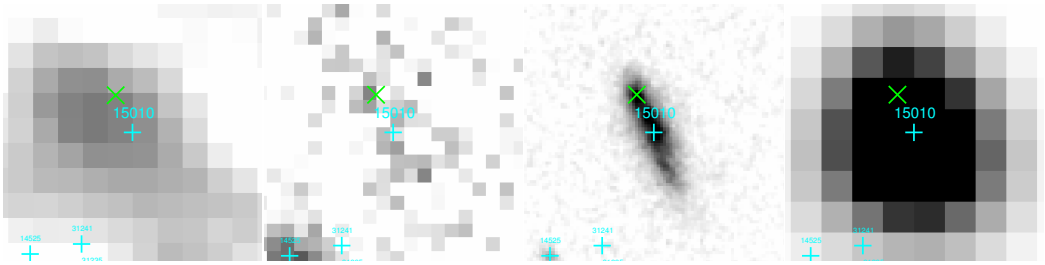
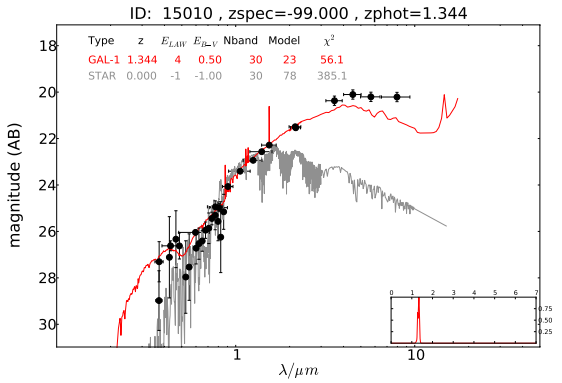


Figure 5.10: Multi-wavelength images of the X-ray source (X-645) with two H-band detections. One with spec-z = 3.47 (source 3372), and the other one (source 3395) with spec-z = 0.665. The red dashed-circle is centered at the X-ray position with the positional error as radius. Cyan crosses indicates the position of H-band association.

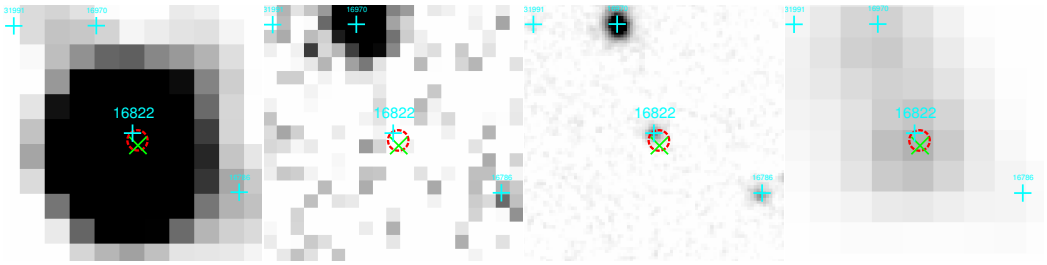


(a) Cutout of M10515 (ID_{H14} 15010)

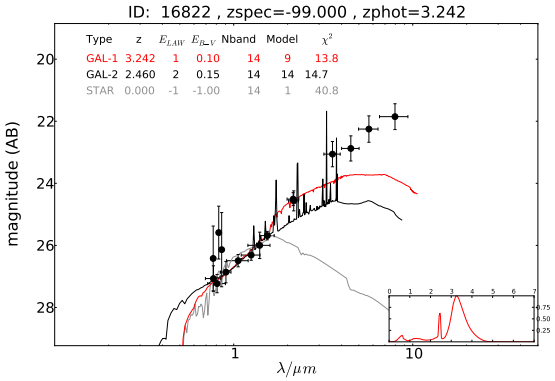


(b) SED of ID_{H14} 15010

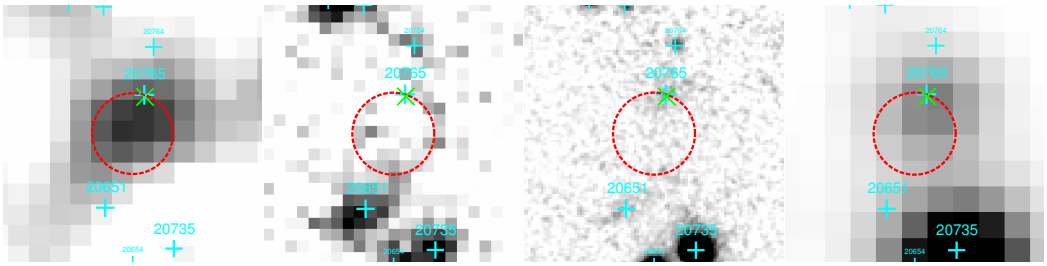
Figure 5.11: Extremely high-z($z > 5$)candidates from Fiore et al. (2012). The cutouts from left to right is in X-ray, R-band, H-band, and IRAC-3.6μm. The red dashed-circle is centered at X-ray position with the positional error on radius. The cyan crosses indicate the positions of H-band associations found in our work, while the green ones indicate the positions of counterparts found by Fiore et al. (2012). We enlarged the font size of the ID number for the likely counterparts.



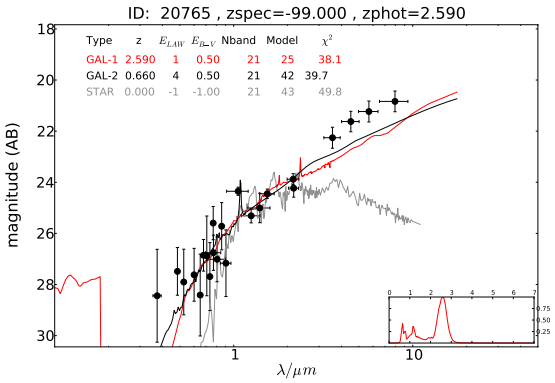
(c) Cutout of M70168 (ID_{H14} 16822)



(d) SED of ID_{H14} 16822

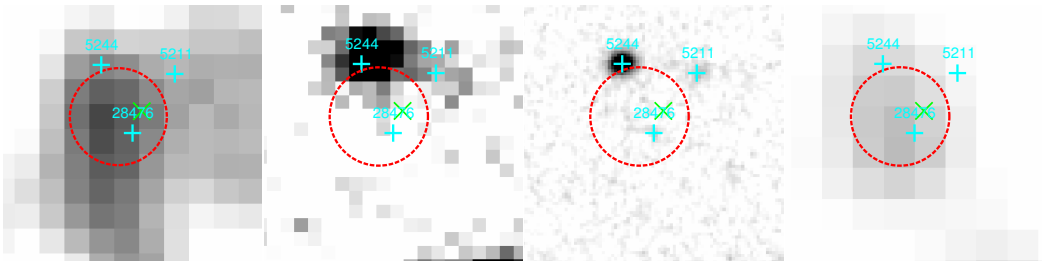


(e) Cutout of E2551 (ID_{H14} 20765). The three H-band associations, 20765, 20735, and 20651 (with posterior p=0.98, 0.82, 0.80, respectively), have photo-z values of 2.59, 1.25, 0.81, correspondingly.

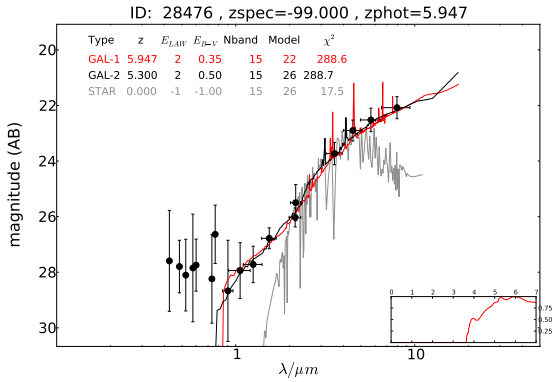


(f) SED of ID_{H14} 20765

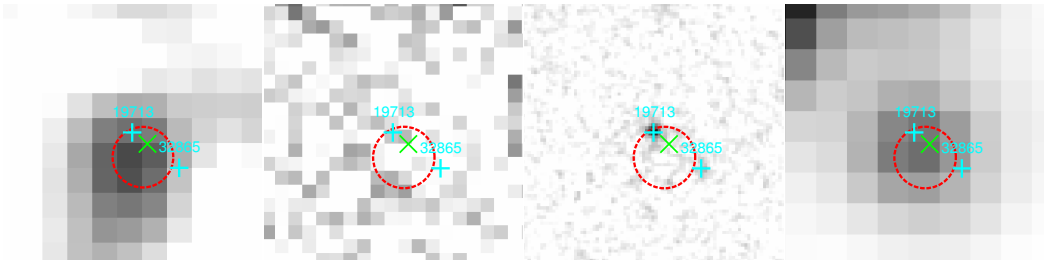
Figure 5.10: (Continued)



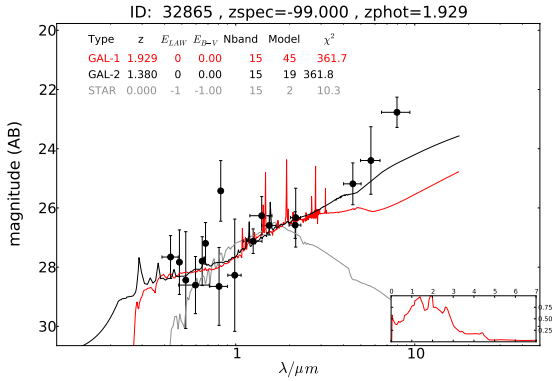
(g) Cutout of M70407 (ID_{H14} 28476). The three H-band associations, 5211, 5244, and 28476 (with posterior p=0.88, 0.95, 0.95, respectively), have photo-z values of 1.66, 2.95, 5.94, correspondingly.



(h) SED of ID_{H14} 28476

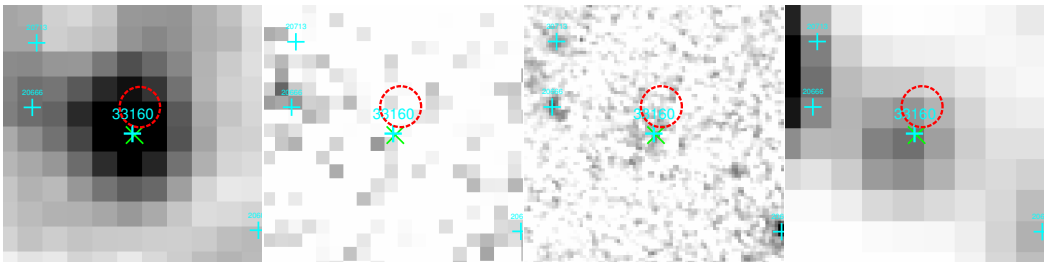


(i) Cutout of M70099 (ID_{H14} 32865). The two H-band associations, 19713 and 32865 (with posterior p=0.99 and 0.80, respectively), have photo-z values of 2.40, and 1.93, correspondingly.

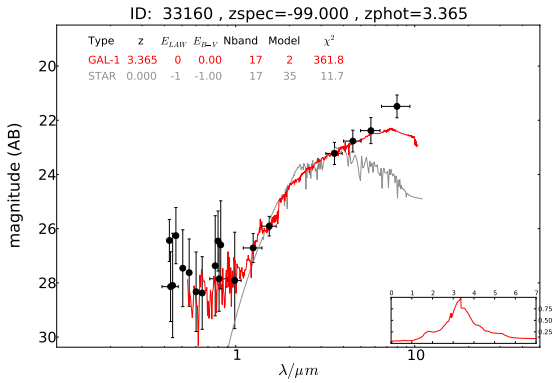


(j) SED of ID_{H14} 32865

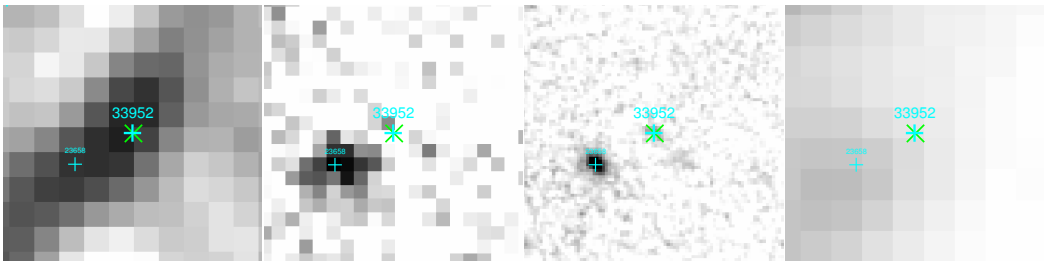
Figure 5.9: (Continued)



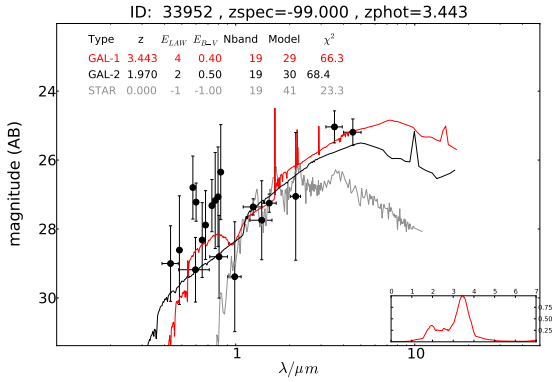
(k) Cutout of E2498 (ID_{H14} 33160)



(l) SED of ID_{H14} 33160



(m) Cutout of E4956 (ID_{H14} 33952)



(n) SED of ID_{H14} 33952

Figure 5.8: (Continued)

Table 5.10: Non-X-ray-detected high-z sources selected by Fiore et al. (2012).

$Name_F$	z_F	$zerr_F$	ID_{H14}	spec-z	z1	zerr1	z2	zerr2	Library
Fiore et al. (2012)					($-24 < M_B < -8$)		($-30 < M_B < -20$)		
M387	3.130	(2.4-3.5)	436	-99	2.6926	(2.46-2.88)	2.6926	(2.46-2.88)	
M603	3.360	(2.2-3.9)	622	-99	3.0130	(0.08-3.34)	3.0130	(2.68-3.34)	
M70531	3.260	(> 1.9)	655	-99	3.0039	(2.93-3.10)	3.0039	(2.93-3.10)	
M70525	3.280	(0.8-3.8)	1000	-99	2.7654	(1.68-3.14)	2.7654	(2.17-3.14)	
M70508	3.180	(2.0-4.0)	1513	-99	1.4621	(1.33-2.08)	1.4621	(1.33-2.08)	
M2690	3.620	(1.5-4.6)	2619	-99	2.1232	(1.93-2.23)	2.1232	(1.93-2.23)	
M70481	4.080	(3.3-4.5)	2717	-99	3.0847	(2.08-3.58)	3.0847	(2.08-3.58)	
M70467	4.250	(> 2.0)	3197	-99	2.8373	(1.92-3.37)	2.8373	(2.37-3.37)	
M3268	3.300	(2.6-3.8)	3260	-99	2.4856	(2.00-3.36)	2.9939	(2.75-3.37)	
M3607	3.468	spec	3768	3.4680	3.3554	(3.25-3.52)	3.3554	(3.25-3.52)	
M4417	3.470	spec	5001	3.4700	3.5690	(3.48-3.65)	3.5690	(3.48-3.65)	
M5522	3.690	(2.7-4.1)	6433	-99	2.9667	(2.68-3.19)	2.9667	(2.68-3.19)	
M7305	3.440	(3.1-3.7)	8932	-99	3.3459	(3.23-3.46)	3.3459	(3.23-3.46)	
M8039	3.190	(2.5-3.9)	10092	-99	3.2593	(2.92-3.52)	3.5742	(3.16-3.81)	
M8728	4.820	(3.9-5.5)	11287	-99	1.0441	(0.66-1.89)	4.4827	(4.21-5.12)	
M9363	3.610	(3.0-4.1)	12380	-99	0.4544	(0.14-3.81)	3.7643	(3.51-4.05)	
M9350	3.360	(3.2-3.5)	12448	-99	3.3400	(3.23-3.40)	3.3400	(3.23-3.40)	Ilbert et al. 2009
M10390	3.210	(2.4-3.7)	14521	-99	1.9270	(1.78-2.07)	1.9270	(1.78-2.07)	
M10548	4.823	spec.	14800	4.8237	4.8146	(4.73-4.92)	4.8146	(4.73-4.92)	
M10515	5.260	(1.8-6.6)	15010	-99	1.3439	(1.22-1.42)	1.3439	(1.22-1.42)	
M13473	3.210	(2.6-3.8)	19706	-99	0.3440	(0.30-0.40)	3.0425	(2.94-3.11)	
E1577	3.462	spec.	19824	3.4739	3.4169	(3.35-3.44)	3.4169	(3.35-3.44)	
E1611	3.356	spec.	19892	3.3560	1.8282	(1.51-2.16)	1.8282	(1.76-2.16)	
E7840	3.680	(3.3-4.0)	24011	-99	2.0990	(1.21-2.61)	2.5742	(2.37-2.76)	
E9850	3.930	(3.8-4.1)	24806	-99	0.9979	(0.19-3.02)	3.3641	(3.12-4.01)	
E10247	3.950	(3.6-4.0)	24833	-99	0.0938	(0.05-0.11)	3.0746	(3.00-3.26)	
E8439	4.910	(4.8-5.2)	25579	-99	1.0551	(0.84-1.37)	5.1486	(4.99-5.46)	
E8479	3.670	(3.3-3.9)	25686	-99	0.2142	(0.04-3.56)	3.3500	(3.14-3.90)	
E7911	4.860	(4.7-5.4)	26030	-99	3.1061	(0.74-3.31)	3.1061	(2.08-3.31)	
E2199	4.710	(4.5-5.2)	33073	-99	0.8887	(0.34-2.18)	4.4233	(4.18-5.25)	
E2658	3.740	(3.5-4.1)	33197	-99	0.4955	(0.10-3.07)	3.5025	(3.26-4.40)	
E4956	7.430	(> 7.0)	33952	-99	3.4428	(1.07-4.99)	4.0949	(3.97-7.00)	
E5165	3.870	(3.6-4.2)	34015	-99	3.3876	(2.40-4.33)	3.9052	(3.64-4.36)	

Note: $Name_F$, z_F, and $zerr_F$ are the redshift informations from Fiore et al. (2012); ID_{H14} are the IDs from our work , with their corresponding spec-z. z1, zerr1 are the photo-z and their 3σ errors computed with magnitude prior: $-24 < M_B < -8$; z2, zerr2 are computed with magnitude prior: $-30 < M_B < -20$; The last column indicates the libraries (already defined in Chapter 3) used to obtained z1, zerr1, z2, and zerr2. The library used for non-X-ray-detected sources consist of normal galaxy templates derived from Ilbert et al. (2009).

Table 5.11: X-ray detected high-z sources selected by Fiore et al. (2012).

$Name_F$	z_F	$zerr_F$	ID_{H14}	spec-z	z1	zerr1	z2	zerr2	Library
Fiore et al. (2012)					$(-24 < M_B < -8)$		$(-30 < M_B < -20)$		
M3320	3.470	spec	3372	3.4700	3.5214	(0.34-3.58)	3.5214	(3.26-3.58)	
M70437	4.420	(> 2.7)	4356	-99	2.4750	(0.18-4.93)	3.0887	(2.89-4.94)	
M70435	3.230	(> 2.0)	4438	-99	2.4463	(1.13-3.41)	2.4463	(2.31-3.41)	
M70429	3.630	(1.2-4.8)	4624	-99	3.1730	(1.87-4.57)	3.1730	(2.57-4.57)	
M70390	4.180	(2.2-4.8)	5652	-99	1.7189	(0.32-3.37)	2.3732	(1.78-3.43)	
M8273	3.064	spec.	10578	3.0640	2.8350	(2.66-3.09)	2.8350	(2.66-3.09)	
M70168	5.010	(2.0-6.5)	16822	-99	3.2419	(0.37-4.53)	3.2419	(2.71-4.53)	
E537	4.290	(2.9-4.5)	18354	-99	2.5959	(2.06-2.92)	2.5959	(2.06-2.92)	
E737	4.140	(4.0-4.4)	18694	-99	0.6080	(0.20-3.32)	3.0432	(2.21-3.52)	Lib-EXT
E1516	4.290	(4.1-6.9)	19713	-99	2.4021	(1.13-4.01)	2.9540	(2.53-4.08)	
E1617	3.380	(3.3-3.4)	19883	-99	2.0952	(0.35-3.66)	2.1447	(2.06-3.66)	
E2551	5.190	(5.1-5.5)	20765	-99	2.5899	(0.53-3.15)	2.5899	(2.05-3.15)	
E2309	3.610	(3.3-3.8)	20936	3.2560	2.9849	(2.40-3.07)	2.9849	(2.40-3.07)	
E6257	3.220	(3.2-3.5)	22426	-99	3.2471	(0.15-3.33)	3.2471	(2.77-3.33)	
M70340	3.900	(> 3.9)	29323	-99	4.6509	(0.19-7.00)	4.6509	(3.37-7.00)	
M70099	6.220	(> 2.9)	32865	-99	1.9287	(0.03-5.71)	4.4047	(4.04-7.00)	
E2498	6.620	(5.9-7.3)	33160	-99	3.3653	(0.04-7.00)	3.3653	(2.70-7.00)	
M208	4.762	spec.	273	4.7600	0.1374	(0.12-4.81)	4.6946	(4.53-4.86)	
M3323	3.471	spec.	3360	3.4710	3.4302	(3.41-3.46)	3.4302	(3.41-3.46)	
M4302	3.610	spec.	4936	3.6100	3.4960	(3.49-3.51)	3.5615	(3.51-3.58)	
M4835	3.660	spec.	5479	3.6600	3.6456	(3.60-3.70)	3.6456	(3.60-3.70)	
M70385	3.980	(1.1-4.7)	5867	-99	2.6819	(1.85-3.09)	2.6819	(2.16-3.09)	Lib-PT
M5390c	3.700	spec.	6294	3.7000	0.1354	(0.11-3.74)	3.6948	(3.65-3.74)	
M5842	3.830	(2.8-4.3)	6905	-99	3.6392	(3.45-3.80)	3.6392	(3.45-3.80)	
M10429	3.193	spec.	14587	3.1930	3.2220	(3.18-3.28)	3.2220	(3.18-3.28)	
E1312	3.480	(3.4-3.6)	19505	-99	3.4165	(0.28-3.95)	3.4165	(3.24-3.95)	
M70407	5.900	(> 1.8)	28476	-99	5.9474	(0.04-7.00)	5.9474	(3.63-7.00)	

Note: $Name_F$, z_F, and $zerr_F$ are the redshift informations from Fiore et al. (2012); ID_{H14} are the IDs from this work , with their corresponding spec-z. z1, zerr1 are the photo-z and their 3σ errors computed with magnitude prior: $-24 < M_B < -8$; z2, zerr2 are computed with magnitude prior: $-30 < M_B < -20$; The last column indicates the libraries (already defined in Chapter 3) used to obtained z1, zerr1, z2, and zerr2. The two libraries used here are tuned for X-ray selected AGNs, inclusive of the hybrid templates constructed for this work as described in Chapter 3.

Table 5.12: High-z ($z > 5$) candidates from Fiore et al. (2012)

$\mathrm{Name_F}$	$\mathrm{z_F}$	$\mathrm{zerr_F}$	$\mathrm{ID_{H14}}$	$\mathrm{ID_{R13}}$	$\mathrm{ID_{X11}}$	z_{H14}	$\mathrm{z_{H14}} \pm 3\sigma$	$w(z \geq 3)$	xflag
M10515	5.26	(1.8-6.6)	15010	-99	-99	1.3439	(1.22-1.42)	0	-99
M70168	5.01	(2.0-6.5)	16822	130	371	3.2419	(0.37-4.53)	0.71	1
E2551	5.19	(5.1-5.5)	20765	550	521	2.5899	(0.53-3.15)	0.02	3
M70407	5.9	(> 1.8)	28476	541	-99	5.9474	(3.63-7.0)	1.0	3
M70099	6.22	(> 2.9)	32865	522	392	1.9287	(0.03-5.71)	0.02	2
E2498	6.62	(5.9-7.3)	33160	444	85	3.3653	(0.04-7.0)	0.69	1
E4956	7.43	(> 7.0)	33952	-99	-99	3.4428	(1.07-4.99)	0.65	-99

Note: $\mathrm{Name_F}$, $\mathrm{z_F}$, and $\mathrm{zerr_F}$ are the redshift informations from Fiore et al. (2012); $\mathrm{ID_{H14}}$, $\mathrm{ID_{R13}}$ and $\mathrm{ID_{X11}}$ are IDs from Hsu et al. (2014); Rangel et al. (2013); Xue et al. (2011), respectively. z_{Hue}, $\mathrm{z_{Hue}} \pm 3\sigma$, and $w(z \geq 3)$ are the photo-z information from Hsu et al. (2014). Non of these sources have spec-z. "xflag" indicates the number of optical/NIR counterparts that were found for the associated X-ray source (done by Hsu et al. 2014). xflag= -99 means that no X-ray association was found.

Table 5.13: Newly-found high-z ($z > 3$) X-ray sources in CANDELS/GOODS-S from Hsu et al. (2014)

$\mathrm{ID_{H14}}$	$\mathrm{ID_{R13}}$	$\mathrm{ID_{X11}}$	zbest	zbest$\pm 3\sigma$	$w(z \geq 3)$
3973	23	460	3.876	(3.05-4.28)	0.98
5064	402	258	3.0503	(2.02-4.0)	0.567
5367	-99	374	3.5187	(0.22-3.76)	0.718
8833	311	132	3.5148	(3.31-3.63)	1.0
9698	408	458	3.4886	(3.25-3.8)	0.99
12309	-99	150	3.4382	(2.77-3.69)	0.905
13327	107	430	3.2405	(2.38-4.02)	0.92
23382	334	534	4.379	spec-z	1.0
24636	486	602	3.6995	(0.23-3.98)	0.884
26315	551	272	3.019	(0.28-4.09)	0.257

$\mathrm{ID_{H14}}$: ID from Hsu et al. (2014);
$\mathrm{ID_{R13}}$: ID from Rangel et al. (2013);
$\mathrm{ID_{X11}}$: ID from Xue et al. (2011);
zbest: spec-z or photo-z derived from Hsu et al. (2014);
zbest$\pm 3\sigma$: 3σ range of the photo-z.
$w(z \geq 3)$: see the definition in Equation 5.4.

5.6 Conclusions

Taking advantage of deep observations from the 4Ms-CDFS Chandra survey, HST/WFC3 NIR survey, and TENIS $J\&K$-band survey that allowed us to improve on the photo-z accuracy, we studied the high-z sources in the GOODS-S region (Area 1). In this chapter, first we investigated the traditional $\mathrm{LBG_{V09}}$, BzK, VJL, iHM selection methods for high-z sources. Then we compared the samples determined in that way and with two additional methods: a simple *zbest* (i.e., considering all sources with certain redshift from either spec-z or photo-z) and the $P(z)$ technique. The latter consider the entire redshift probability distribution function and the area beneath it within the redshift range of interest. We compare the methods against each by assessing their efficiency expressed as completeness and purity. We conclude the main results as following:

- Comparing with $\mathrm{LBG_{V09}}$, BzK, VJL, and iHM selections, $P(z)$-selected samples have much higher values of $(C + P_s)$ it is the most efficient and reliable method to select high-z sources. This is because the $P(z)$-technique makes use of photometric information from the entire SED rather than using only three photometric points. In addition, the use of $P(z)$ allows us to modify when necessary, the threshold w to create more pure subsamples.

- We investigated the efficiencies of color methods and found that the color criteria are not perfectly applicable for selecting high-z AGNs from our X-ray sample. For example, at $z \sim 3$, comparing tje $\mathrm{LBG_{V09}}$ and VJL color methods, we find that they have similar completeness (70%) but the former is much more contaminated by low redshift sources (70% purity for the VJL against the 30% of the $\mathrm{LBG_{V09}}$). The reason for such low efficiency is that color methods are originally tuned for normal galaxies and the different SEDs of AGNs effect the selection making the samples less complete and pure.

- Comparing the efficiencies of color selections for high-z AGNs and normal galaxies, we found that the AGN samples generally reach higher completeness and purities than the normal galaxy samples. This indicates that the color selections discussed in our work are more efficient to select high-z AGNs than normal galaxies.

Second, applying the $P(z)$ technique, we built a high-z ($z > 3$) list for X-ray sources in ECDFS region. We compared our list with previous work done by Fiore et al. (2012) in GOODS-S region. We found some discrepancies because of:

- Different optical/NIR counterparts are assigned. In our work, we made better X-ray-to-optical/NIR associations with better consideration on the positional errors and magnitude distribution.

- Photo-z results could change with different magnitude priors and improper templates for SED-fitting. We used well-established AGN-galaxy hybrids for X-ray selected

AGNs and applied magnitude priors on point-like and extended sources. These procedures help us reduce the degeneration and improve our photo-z accuracy.

- Uncertainty of the photo-z may become large due to the limitation of the optical/IR observation. Sources at high redshift usually have faint photometry and broad $P(z)$. This is the reason why we should take into account the $P(z)$ when using photo-z, in particular at high redshift.

- From our high-z sample in GOODS-S region, we found ten more high-z X-ray sources at $z > 3$ with reliable photo-z.

Chapter 6

Summary and prospects for the future

To better investigate the AGN and galaxies coevolution, we need to have a complete characterization and census of AGN with accurate redshift. This can be obtained by (1) defining properly templates for AGNs in order to properly compute reliable photo-z; (2) studying the colors of AGN host, comparing them with the properties of normal galaxies; (3) comparing number density of AGN and galaxies selected at high redshift. In this thesis, we have fully accomplished the first two items. The work was done in the field centered in the ECDFS field including the CANDELS/GOODS-S and CDFS areas, where the deepest optical/NIR and X-ray data are existing and thus a better sampling can be obtained. With these good-quality data, for X-ray sources, we achieved a photo-z accuracy of ~ 0.013 with an outliers fraction of $\sim 5.3\%$. For normal galaxies, the photo-z accuracy is ~ 0.010 with an outlier fraction of $\sim 4.6\%$ (see Chapter 3). Following we summarize the steps for computing photo-z:

- Due to the low resolution of X-ray data, a simple match in coordinates is not enough to provide confident counterparts for X-ray sources, specially if the data are as deep as those provided by the CANDELS. Taking into account positional errors and multiple magnitude distributions as priors simultaneously, we adopted Bayesian statistics for determining the most likely counterparts to the X-ray sources from literature. We cross-matched the CANDELS, MUSYC, TENIS, and IRAC catalogs all together with the X-ray catalogs from the 250ks-ECDFS survey (Lehmer et al., 2005a; Virani et al., 2006) and the 4Ms-CDFS survey (Xue et al., 2011; Rangel et al., 2013). We found that more then 96% of X-ray sources in the CDFS/ECDFS have optical/NIR counterparts, 21% of these counterparts are not unique (more than one counterpart are found for a X-ray source) (Chapter 2).

- In addition to the space-based CANDELS dataset, we added the deepest ground-based NIR data from the TENIS survey, intermediate-band photometry from the MUSYC survey, and UV data form the *GALEX*. The *GALEX* data, despite the shallowness, allow us to characterize the UV excess (which is the typical feature of

the type 1 AGN or star formation), limiting the fraction of photo-z outliers. Equally important, the intermediate-band data help in identifying the strong emission line features in AGN SEDs, thus enabling more precise photometric redshifts for both AGN and normal galaxies. Correctly merging these multi-wavelength catalogs is an essential step before our photo-z computation. To do this, we first removed the positional offsets between catalogs, then we calibrated flux difference between photometries (total fluxes, flux apertures or PSF-fitted photometry) (Chapter 2).

- For galaxies, the templates used to fit SED and compute photo-z have been well established in the past few years. However for AGN, the SED consists of mixed radiation emitted from the central AGN and the host galaxy, which can not be represented by pure galaxy or pure AGN template. We improved on previous SEDs by creating new hybrids with more realistic templates of normal galaxies by (1) taking into account the nature of AGNs where emission lines are always present, and (2) considering that emission lines are also emitted from the host galaxies in some cases and contribute to SED together with emission lines from AGNs at high redshift. The hybrid templates were built using the galaxy templates from Bender et al. (2001); Noll et al. (2004) in which empirical emission lines are considered, combining with Type 1 and Type 2 QSO from Polletta et al. (2007). Moreover, we considered morphology (extended/ point-like sources) in order to reduce the degeneracy via an absolute magnitude priors. We used 25% of X-ray sources with spec-z in the CDFS (mainly consists of faint population) to tune the hybrids for point-like and extended sources separately, and applied the best selection of templates to all the other X-ray sources in 4Ms-CDFS regions. For bright point-like AGN population in 250-ks ECDFS survey, we adopted hybrids directly from Salvato et al. (2009) who tuned the hybrids using brighter population from the shallower COSMOS survey (Chapter 3).

Secondly we have investigated the similarities between AGN hosts and normal galaxies by mean of the color-magnitude diagram. We found that as for galaxies, AGN host colors present a bimodality in the CMD up to $z \sim 2.5$ after the correction for dust extinction and AGN contribution. Compared the global color distributions between AGN hosts and galaxies, the blue groups of AGN samples are redder than the blue cloud in the galaxies, while the red groups of AGN samples are bluer than the red sequence in the galaxies. In addition, the positions of blue peaks in the AGN samples are almost constant with cosmic time. This implies a weak connection between AGN activity and star formation in their host. For the X-ray sources in the 4Ms-CDFS survey, we found that for most of the sources, the correction for dust extinction is larger than the correction for the AGN contribution. This is because the AGN population in this field is dominated by low-luminosity type 2 AGNs which have host-dominated SEDs. However for few bright sources, their host colors are strongly effected by AGN contribution rather than by the dust extinction. For these sources the correction for AGN contribution is about two times larger than the correction for dust extinction in general. Therefore a proper AGN/galaxy decomposition becomes

crucial when studying the AGN host properties in the shallower and wider X-ray surveys, e.g., XMM-COSMOS, Stripe82X, XMM-XXL, and more importantly for the future eROSITA all-sky survey, which are biased toward bright AGNs. For sources in these fields, ignoring the AGN contribution would invalidate any result derived by studying their CMD distribution(Chapter 4).

With the reliable optical/NIR counterparts and improved photo-z , we investigated the selection methods for both high-z AGNs and galaxies, including the $\mathrm{LBG_{V09}}$, BzK, VJL, iHM, and the $P(z)$ techniques. We quantify the efficiency of various techniques by comparing the maximum values ($C+P_s$) (i.e. completeness and purity) taking into account the redshift probability distribution function $P(z)$. For both galaxies and AGNs, the $P(z)$ technique is much more reliable than other color methods, as the information from entire SED are used rather than using only two or three bands as it is done in the color selection techniques. Part of the inefficiency that we found can be due to the fact that the traditional color methods are originally tuned for normal galaxies, using specific filters and requiring certain observational depth. In fact, we have used different set of filters, on different depth of data and searching for different type of sources (AGN rather than galaxies). We found that depending on the types of sources the reliability of the methods can be improved by slightly changing the range of redshift where the methods where supposed to be efficient (Chapter 5).

Finally, we built a high-z ($z > 3$) sources list for the X-ray sources in the ECDFS region, and compared our list with previous work done by Fiore et al. (2012) in the GOODS-S region. We found that the discrepancies are caused mainly by the totally different approach in selecting the sample. We selected the most likely counterparts to the X-ray sources depending on their properties, defining proper templates and priors to be used, and then considered the entire $P(z)$ for each counterpart. Instead Fiore et al. (2012) started from a sample of optical/NIR sources with high spec-z or best-fit photo-z, searching for the X-ray detections at those optical/NIR positions. We demonstrated that the latter approach is very unstable and produce incomplete and less pure samples (Chapter 5).

The latter exercise reveals the problem of identifying counterparts which will be more severer for the next all-sky X-ray surveys planned with the eROSITA. Namely, if with the already high positional accuracy and image resolution of *Chandra*, the association to a multi-wavelength counterpart is still problematic in certain cases, it will be even more difficult for the sources detected with the shallower eROSITA, given its lower resolution ($\sim 20''$). For future counterpart catalogs of the eROSITA sources, one must carry along the posterior value of the associated counterpart. Moreover, when studying AGN/galaxy evolution at the bright end and high redshift regime, instead of using best-fit photo-z, the entire $P(z)$ must be considered.

Future perspective

With the accurate photo-z we obtained and the AGN/galaxy decomposition we accomplished, we can extend the researches of AGN and galaxy (co)evolution to more distant

Universe. The next step of researches we are going to approach is described below:

- **AGN duty cycle.** One way to investigate AGN evolution is to understand the AGN duty cycle (i.e., the time scale that SMBH is active) by estimating AGN fraction among the normal galaxies. Utilizing the deep *Chandra* and *HST* observations, we can have more complete X-ray-selected AGN demography to better constrain AGN population at the faint end of the luminosity function. However we know that a large fraction of galaxies that host AGNs do not emit detectable X-rays but are identified at optical, infrared and/or radio wavelengths. To achieve a more complete census of AGN to investigate the AGN and galaxies (co)evolution, it is necessary to combine AGNs from various depth of multi-wavelength surveys from X-ray to radio (including imaging and spectroscopy), e.g., the NIR observation from the WISE or VISTA surveys, and the optical data from the HSC or SDSS-IV survey. Additionally, the future all-sky eROSITA surveys is expected to detect millions of AGN in soft and hard X-ray. With the correctly determined X-ray counterparts, The combination of multi-band surveys will help in completing the framework of the AGN evolution.

- **Stellar masses of AGNs.** Studying the stellar masses of AGN can provide links between the central SMBH and their host galaxies, such as AGN feedback on the host star-forming process. However subtracting the AGN contribution from the SED is always a big challenge, specially for luminous AGN-dominated sources at higher redshift. We are planning to extend the AGN/host decomposition we did for the color-magnitude relation to the host stellar mass computation, also taking into account the degeneracies in the SED fitting and a proper error budged. The final goal is to analyze the AGN host mass function, and star formation rate, etc, to understand the AGN Host properties.

Appendix A

Released Catalogs

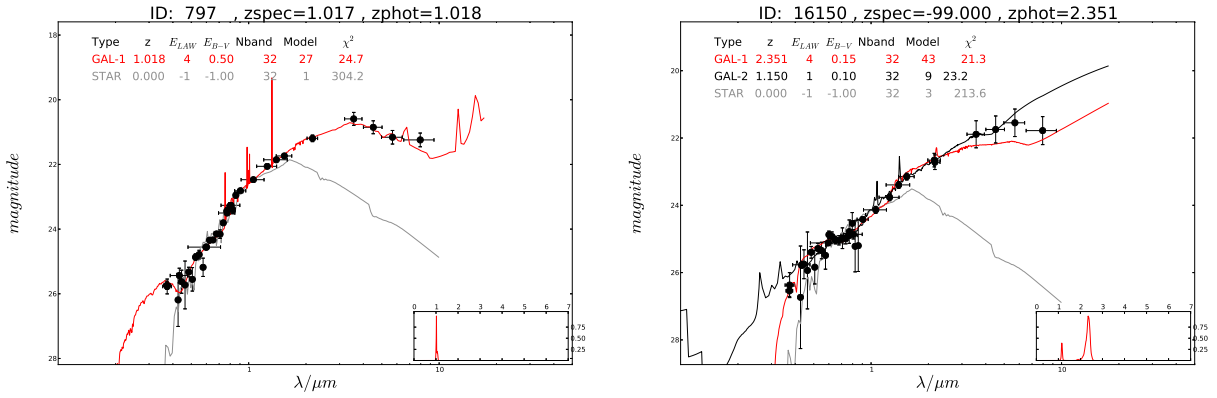
Tables 8 through 11 give homogeneously computed photo-z and related data for all sources detected in the area covered by CANDELS/GOODS-S, CDFS, and ECDFS survey. For each source we make also available the redshift probability distribution function $P(z)$[1]. With these data it is possible to construct figures like the inserts in Figure A.1. Because of the large size of the $P(z)$ files, we provide them at the link `http://www.mpe.mpg.de/XraySurveys`. In lieu of the full $P(z)$, the catalogs provide a proxy in the form of the normalized integral of the main probability distribution $P(z_p) \equiv 100 \times \int P(z)dz$ with the integral over the range $z_p \pm 0.1(1 + z_p)$. A value close to 100 indicates that the photo-z value is uniquely defined. Smaller values imply that a wide range or multiple photo-z values are possible.

For the *Chandra* X-ray detections, the catalogs also provide a new compilation of X-ray source lists from the literature, the new optical/NIR/MIR associations, and the corresponding photometry. Catalog descriptions and excerpts are below. An entry of -99 indicates no data for that quantity. All coordinates are J2000. Updated versions of the catalogs and templates will be available under [Surveys] > [CDFS] through the portal `http://www.mpe.mpg.de/XraySurveys` .

A.1 Cross ID reference table

Table A.1 gives cross-IDs and positions for all sources within each Area as identified in Table 2.3. The table also indicates whether a source is a possible counterpart to an X-ray detection.

[1]The redshift probability distribution function is derived directly from the χ^2: $P(z) \propto \exp\left(-\frac{\chi^2(z)-\chi^2_{\rm min}}{2}\right)$; 1σ is estimated from $\chi^2(z)-\chi^2_{\rm min} = 1$ (68%), and 2.3σ is estimated from $\chi^2(z)-\chi^2_{\rm min} = 6.63$ (99%)



(a) (b)

Figure A.1: Two examples of SED fitting for source 797 (a normal galaxy) and source 16150 (a X-ray-detected AGN). The photometric points are shown in black. The red lines show the best-fitting template, and grey lines the best fitting star (the latter a poor fit for both objects shown). In the right panel, the black line shows the second-best template. Information about the templates—type, photo-z, extinction law, extinction value, number of bands, model identification, and χ^2 of the fit—is given in the legends. Inserts show $P(z)$ for the sources.

Table A.1: Column description of the cross ID reference catalog

Column	Title	Description
1	[HSN2014]	Sequential number adopted in this work.
2-4	ID_C , $R.A._C$, $Dec._C$	ID, right ascension and declination from the CANDELS-TFIT catalog (G13).
5-7	ID_M , $R.A._M$, $Dec._M$	ID, right ascension and declination from the MUSYC catalog (Cardamone et al., 2010b).
8-10	ID_T, $R.A._T$, $Dec._T$	ID, right ascension and declination from the TENIS catalog (Hsieh et al., 2012).
11-13	ID_S, $R.A._S$, $Dec._S$	ID, right ascension and declination from the SIMPLE catalog (Damen et al., 2011).
14-17	ID_{R13}, $R.A._{R13}$, $Dec._{R13}$, $PosErr_{R13}$	ID, right ascension, declination and positional error from the R13 4Ms-CDFS catalog.
18-21	ID_{X11}, $R.A._{X11}$, $Dec._{X11}$, $PosErr_{X11}$	ID, right ascension, declination and positional error from the X11 4Ms-CDFS catalog.
22-25	ID_{L05}, $R.A._{L05}$, $Dec._{L05}$, $PosErr_{L05}$	ID, right ascension, declination and positional error from the L05 250ks-ECDFS catalog.
26-29	ID_{V06}, $R.A._{V06}$, $Dec._{V06}$, $PosErr_{V06}$	ID, right ascension, declination and positional error from the V06 250ks-ECDFS catalog.
30	Xflag	"1" indicates that the source is the only possible counterpart to an X-ray source. "n" (2 or more) indicates that the source is one of the "n" possible counterparts for a give X-ray source. "-99" indicates that no X-ray counterpart are found.
31	p	Posterior value which indicates the reliability of the X-ray to optical/NIR/MIR association. (as defined in Section 2.2.2)

Table A.2: X-ray source list

[HSN2014] (1)	ID_{R13} (2)	ID_{X11} (3)	ID_{L05} (4)	ID_{V06} (5)	R.A.$_x$ (6)	DEC.$_x$ (7)	$Flux_s$ (8)	$Flux_h$ (9)	$Flux_f$ (10)	log L_s (11)	log L_h (12)	log L_f (13)
125	343	266	-99	-99	53.079439	-27.949429	2.05E-16	7.62E-16	9.86E-16	41.28	41.85	41.969
482	6	336	-99	-99	53.103424	-27.933357	8.84E-16	3.67E-15	4.59E-15	42.37	42.99	43.085
47821	-99	-99	527	445	53.251375	-27.980556	1.06E-15	2.33E-15	3.22E-15	42.66	43.00	43.14
50721	-99	-99	32	348	52.842417	-27.965417	2.07E-15	1.61E-14	1.81E-14	42.14	43.03	43.08

A.2 X-ray source list in ECDFS

Table A.2 gives the X-ray source list in Areas 1+2+3 with the position and flux information from the available catalogs. Columns are as follows:
(1) [HSN2014]: Sequential number adopted in this work.
(2) ID_{R13}: ID from R13 catalog
(3) ID_{X11}: ID from X11 catalog
(4) ID_{L05}: ID from L05 catalog
(5) ID_{V06}: ID from V06 catalog
(6) R.A.$_x$: Right Ascension of the X-ray source.
(7) DEC.$_x$: Declination of the X-ray sources.
(8) $Flux_s$: Soft band X-ray flux (erg cm^{-2} s^{-1}).
(9) $Flux_h$: Hard band X-ray flux.
(10) $Flux_f$: Full band X-rayflux.
(11) log L_s: Soft band X-ray luminosity (erg s^{-1}).
(12) log L_h: Hard band X-ray luminosity.
(13) log L_f: Full band X-ray luminosity.
Note: From column (6) to (10), we chose the original X-ray data from, in order of priority, R13, X11, L05 and V06.

A.3 Photometry of X-ray sources

Table A.3 gives photometry for all the possible counterparts to the X-ray sources. For the CANDELS area, this includes the TFIT photometry in the IBs as described in Section 2.1.2. Columns are as follows:
(1) [HSN2014]: Sequential number adopted in this work.
(2)-(5) XID: ID from the four X-ray catalogs with the same order as Table A.2
(6)Xflag: As described in Table A.1
(7)p: As described in Table A.1
(8)R.A.$_{opt}$: Right Ascension of the optical/NIR/MIR source.
(9)Dec.$_{opt}$: Declination of the optical/NIR/MIR source.
(10)-(109): AB magnitude and the associated uncertainty in each of possible bands (see the detailed list in Table 2.1 and Table 2.2).

Table A.3: Photometry of X-ray sources

[HSN2014] (1)	XID (2)-(5)	xflag (6)	p (7)	R.A.$_{opt}$ (8)	Dec.$_{opt}$ (9)	FUV_m (10)	FUV_e (11)	NUV_m (12)	NUV_e (13)	... ...	IRAC4$_m$ (108)	IRAC4$_e$ (109)
125	...	2	0.98	53.079489	-27.948735	-99.0	-99.0	-99.0	-99.0	... ...	19.888	0.016
482	...	1	0.99	53.103520	-27.933323	-99.0	-99.0	-99.0	-99.0	... ...	21.096	0.03
47821	...	2	0.97	53.252067	-27.980645	-99.0	-99.0	-99.0	-99.0	... ...	22.421	0.18
50721	...	2	1.0	52.84249	-27.965261	-99.0	-99.0	-99.0	-99.0	... ...	19.24	0.032

A.4 Redshift catalog

Table A.4 gives photo-z results for all sources detected in the CANDELS/GOODS-S, CDFS and ECDFS area. X-ray detections are flagged in the catalog. Columns are:
(1) [HSN2014]: Sequential number adopted in this work.
(2) R.A.$_{opt}$: Right Ascension of the optical/NIR/MIR source.
(3) Dec.$_{opt}$: Declination of the optical/NIR/MIR source.
(4) z_s: Spectroscopic redshift (N. Hathi, private communication).
(5) Q_{zs}: Quality of the spectroscopic redshift. (0=High, 1=Good, 2=Intermediate, 3=Poor).
(6) z_p: The photo-z value as defined by the minimum of χ^2.
(7) $1\sigma^{low}$: Upper 1σ value of the photo-z.
(8) $1\sigma^{up}$: Lower 1σ value of the photo-z.
(9) $3\sigma^{low}$: Upper 3σ value of the photo-z.
(10) $3\sigma^{up}$: Lower 3σ value of the photo-z.
(11) $P(z_p)$: Normalized area under the curve $P(z)$, computed between $z_p \pm 0.1(1+z_p)$.
(12) z_p2: The second solution in the photo-z, when the $P(z_{p2})$ is above 5.
(13) $P(z_{p2})$: Normalized area under the curve $P(z)$, computed between $z_{p2} \pm 0.1(1+z_{p2})$.
(14) N_p: Number of photometric points used in the fit.
(15) Mod: Template number used for SED fitting. 1-48 are the templates from Lib-EXT; 101-130 are the templates from Lib-PT; 201-230 are the templates from S09; 301-331 are the templates from Ilbert et al. (2009), in the same order as the mentioned authors used.
(16) Xflag: As described in Table A.1
(17) p: Posterior value which indicates the reliability of the X-ray to Optical/NIR/MIR association (as defined in Section 2.2.2).

Table A.4: Redshift catalogs

[HSN2014] (1)	R.A.$_{\rm opt}$ (2)	Dec.$_{\rm opt}$ (3)	z_s (4)	$Q_{\rm zs}$ (5)	z_p (6)	$1\sigma^{\rm low}$ (7)	$1\sigma^{\rm up}$ (8)	$3\sigma^{\rm low}$ (9)	$3\sigma^{\rm up}$ (10)	$P(z_{\rm p})$ (11)	$z_{\rm p}2$ (12)	$P(z_{\rm p}2)$ (13)	$N_{\rm p}$ (14)	Mod (15)	xflag (16)	p (17)
13	53.093452	-27.957135	-99.0	-99	3.2619	3.25	3.27	3.21	3.29	100.0	-99.0	0.0	29	328	-99	-99
14	53.104490	-27.957068	-99.0	-99	2.1768	0.44	2.23	0.43	2.46	91.05	0.45	8.91	27	331	-99	-99
15	53.088446	-27.956996	-99.0	-99	3.0468	3.01	3.09	2.92	3.18	100.0	-99.0	0.0	26	322	-99	-99
16	53.104181	-27.956592	-99.0	-99	3.1233	3.05	3.19	2.93	3.28	99.99	-99.0	0.0	27	324	-99	-99
125	53.079490	-27.94874	0.619	0	0.6664	0.66	0.68	0.65	0.68	100.0	-99.0	0.0	24	028	2	0.98
135	53.142288	-27.94447	-99.0	-99	2.6422	2.5	2.72	1.11	3.07	66.19	1.17	13.80	27	014	1	0.95

Bibliography

Aird, J., Nandra, K., Georgakakis, A., et al. 2008, MNRAS, 387, 883

Aird, J., Nandra, K., Laird, E. S., et al. 2010, MNRAS, 401, 2531

Aird, J., Coil, A. L., Moustakas, J., et al. 2012, ApJ, 746, 90

Alexander, D. M., & Hickox, R. C. 2012, New A Rev., 56, 93

Alexander, D. M., Bauer, F. E., Brandt, W. N., et al. 2003, AJ, 126, 539

Antonucci, R. 1993, ARA&A, 31, 473

Appenzeller, I., Bender, R., Böhm, A., et al. 2004, The Messenger, 116, 18

Arnouts, S., Cristiani, S., Moscardini, L., et al. 1999, MNRAS, 310, 540

Arnouts, S., Moscardini, L., Vanzella, E., et al. 2002, MNRAS, 329, 355

Ashby, M. L. N., Willner, S. P., Fazio, G. G., et al. 2013, ApJ, 769, 80

Bell, E. F., Wolf, C., Meisenheimer, K., et al. 2004, ApJ, 608, 752

Bender, R., Appenzeller, I., Böhm, A., et al. 2001, in Deep Fields, ed. S. Cristiani, A. Renzini, & R. E. Williams, 96

Benítez, N., Moles, M., Aguerri, J. A. L., et al. 2009, ApJ, 692, L5

Best, P. N., Kauffmann, G., Heckman, T. M., et al. 2005, MNRAS, 362, 25

Bongiorno, A., Merloni, A., Brusa, M., et al. 2012, MNRAS, 427, 3103

Boyle, B. J., Shanks, T., Croom, S. M., et al. 2000, MNRAS, 317, 1014

Boyle, B. J., Shanks, T., & Peterson, B. A. 1988, MNRAS, 235, 935

Brammer, G. B., Whitaker, K. E., van Dokkum, P. G., et al. 2009, ApJ, 706, L173

Brandt, W. N., & Hasinger, G. 2005, ARA&A, 43, 827

Brusa, M., Comastri, A., Daddi, E., et al. 2005, A&A, 432, 69

Brusa, M., Zamorani, G., Comastri, A., et al. 2007, ApJS, 172, 353

Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000

Buchner, J., Georgakakis, A., Nandra, K., et al. 2015, ApJ, 802, 89

Budavári, T., & Szalay, A. S. 2008, ApJ, 679, 301

Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682

Cappelluti, N., Brusa, M., Hasinger, G., et al. 2009, A&A, 497, 635

Cardamone, C. N., Urry, C. M., Schawinski, K., et al. 2010a, ApJ, 721, L38

Cardamone, C. N., Urry, C. M., Damen, M., et al. 2008, ApJ, 680, 130

Cardamone, C. N., van Dokkum, P. G., Urry, C. M., et al. 2010b, ApJS, 189, 270

Castellano, M., Salimbeni, S., Trevese, D., et al. 2007, ApJ, 671, 1497

Civano, F., Elvis, M., Brusa, M., et al. 2012, ApJS, 201, 30

Comastri, A. 2004, in Astrophysics and Space Science Library, Vol. 308, Supermassive Black Holes in the Distant Universe, ed. A. J. Barger, 245

Cowie, L. L., & Barger, A. J. 2008, ApJ, 686, 72

Cowie, L. L., Barger, A. J., Bautz, M. W., Brandt, W. N., & Garmire, G. P. 2003, ApJ, 584, L57

Daddi, E., Cimatti, A., Renzini, A., et al. 2004, ApJ, 617, 746

Daddi, E., Dickinson, M., Morrison, G., et al. 2007, ApJ, 670, 156

Dahlen, T., Mobasher, B., Dickinson, M., et al. 2010, ApJ, 724, 425

Dahlen, T., Mobasher, B., Faber, S. M., et al. 2013, ApJ, 775, 93

Damen, M., Labbé, I., van Dokkum, P. G., et al. 2011, ApJ, 727, 1

Dehghan, S., & Johnston-Hollitt, M. 2014, AJ, 147, 52

Donley, J. L., Koekemoer, A. M., Brusa, M., et al. 2012, ApJ, 748, 142

D'Onofrio, M., Marziani, P., & Sulentic, J. W. 2012, Fifty Years of Quasars: From Early Observations and Ideas to Future Research, 386, doi:10.1007/978-3-642-27564-7

Drory, N., Salvato, M., Gabasch, A., et al. 2005, ApJ, 619, L131

Elitzur, M. 2008, New A Rev., 52, 274

Elston, R., Rieke, G. H., & Rieke, M. J. 1988, ApJ, 331, L77

Elvis, M., Wilkes, B. J., McDowell, J. C., et al. 1994, ApJS, 95, 1

Elvis, M., Civano, F., Vignali, C., et al. 2009, ApJS, 184, 158

Fan, X., Strauss, M. A., Schneider, D. P., et al. 2001, AJ, 121, 54

Fan, X., Strauss, M. A., Richards, G. T., et al. 2006, AJ, 131, 1203

Feulner, G., Gabasch, A., Salvato, M., et al. 2005, ApJ, 633, L9

Fiore, F., Puccetti, S., Brusa, M., et al. 2009, ApJ, 693, 447

Fiore, F., Puccetti, S., Grazian, A., et al. 2012, A&A, 537, A16

Ford, H. C., Harms, R. J., Tsvetanov, Z. I., et al. 1994, ApJ, 435, L27

Fotopoulou, S., Salvato, M., Hasinger, G., et al. 2012, ApJS, 198, 1

Franx, M., Labbé, I., Rudnick, G., et al. 2003, ApJ, 587, L79

Gabasch, A., Bender, R., Seitz, S., et al. 2004, A&A, 421, 41

Gabasch, A., Hopp, U., Feulner, G., et al. 2006, A&A, 448, 101

Gebhardt, K., Bender, R., Bower, G., et al. 2000, ApJ, 539, L13

Ghez, A. M., Duchêne, G., Matthews, K., et al. 2003, ApJ, 586, L127

Giacconi, R., Zirm, A., Wang, J., et al. 2002, ApJS, 139, 369

Giavalisco, M., Ferguson, H. C., Koekemoer, A. M., et al. 2004a, ApJ, 600, L93

Giavalisco, M., Dickinson, M., Ferguson, H. C., et al. 2004b, ApJ, 600, L103

Goulding, A. D., Alexander, D. M., Lehmer, B. D., & Mullaney, J. R. 2010, MNRAS, 406, 597

Grazian, A., Fontana, A., de Santis, C., et al. 2006a, A&A, 449, 951

—. 2006b, A&A, 449, 951

Grazian, A., Castellano, M., Koekemoer, A. M., et al. 2011, A&A, 532, A33

Grogin, N. A., Kocevski, D. D., Faber, S. M., et al. 2011, ApJS, 197, 35

Guo, Y., Ferguson, H. C., Giavalisco, M., et al. 2013, ApJS, 207, 24

Haas, M., Klaas, U., Müller, S. A. H., et al. 2003, A&A, 402, 87

Hasinger, G. 2008, A&A, 490, 905

Hasinger, G., Miyaji, T., & Schmidt, M. 2005, A&A, 441, 417

Heckman, T. M., Kauffmann, G., Brinchmann, J., et al. 2004, ApJ, 613, 109

Hickox, R. C., Jones, C., Forman, W. R., et al. 2009, ApJ, 696, 891

Hsieh, B.-C., Wang, W.-H., Hsieh, C.-C., et al. 2012, ApJS, 203, 23

Hsu, L.-T., Salvato, M., Nandra, K., et al. 2014, ApJ, 796, 60

Ilbert, O., Arnouts, S., McCracken, H. J., et al. 2006, A&A, 457, 841

Ilbert, O., Capak, P., Salvato, M., et al. 2009, ApJ, 690, 1236

Kennicutt, Jr., R. C. 1998, ARA&A, 36, 189

King, A. 2003, ApJ, 596, L27

Koekemoer, A. M., Alexander, D. M., Bauer, F. E., et al. 2004, ApJ, 600, L123

Koekemoer, A. M., Faber, S. M., Ferguson, H. C., et al. 2011, ApJS, 197, 36

Koratkar, A., & Blaes, O. 1999, PASP, 111, 1

Kormendy, J., & Richstone, D. 1995, ARA&A, 33, 581

Krolik, J. H. 2001, ApJ, 551, 72

Laidler, V. G., Papovich, C., Grogin, N. A., et al. 2007, PASP, 119, 1325

Laird, E. S., Nandra, K., Adelberger, K. L., Steidel, C. C., & Reddy, N. A. 2005, MNRAS, 359, 47

Laird, E. S., Nandra, K., Georgakakis, A., et al. 2009, ApJS, 180, 102

Lehmer, B. D., Brandt, W. N., Alexander, D. M., et al. 2005a, ApJS, 161, 21

—. 2005b, AJ, 129, 1

Luo, B., Bauer, F. E., Brandt, W. N., et al. 2008, ApJS, 179, 19

Luo, B., Brandt, W. N., Xue, Y. Q., et al. 2010, ApJS, 187, 560

Madau, P., & Dickinson, M. 2014, ARA&A, 52, 415

Magorrian, J., Tremaine, S., Richstone, D., et al. 1998, AJ, 115, 2285

Maraston, C. 2005, MNRAS, 362, 799

Marconi, A., & Hunt, L. K. 2003, ApJ, 589, L21

Marconi, A., Risaliti, G., Gilli, R., et al. 2004, MNRAS, 351, 169

Martínez-Sansigre, A., Rawlings, S., Lacy, M., et al. 2006, MNRAS, 370, 1479

Matute, I., Márquez, I., Masegosa, J., et al. 2012, A&A, 542, A20

McCall, M. L., Rybski, P. M., & Shields, G. A. 1985, ApJS, 57, 1

Merloni, A., Rudnick, G., & Di Matteo, T. 2004, MNRAS, 354, L37

Merloni, A., Predehl, P., Becker, W., et al. 2012, ArXiv e-prints, arXiv:1209.3114

Mobasher, B., Dahlen, T., Ferguson, H. C., et al. 2015, ApJ, 808, 101

Moriondo, G., Cimatti, A., & Daddi, E. 2000, A&A, 364, 26

Mouhcine, M., Lewis, I., Jones, B., et al. 2005, MNRAS, 362, 1143

Moustakas, J., Kennicutt, Jr., R. C., & Tremonti, C. A. 2006, ApJ, 642, 775

Nandra, K., George, I. M., Mushotzky, R. F., Turner, T. J., & Yaqoob, T. 1997, ApJ, 477, 602

Nandra, K., Laird, E. S., & Steidel, C. C. 2005, MNRAS, 360, L39

Nandra, K., Georgakakis, A., Willmer, C. N. A., et al. 2007, ApJ, 660, L11

Naylor, T., Broos, P. S., & Feigelson, E. D. 2013, ApJS, 209, 30

Nenkova, M., Ivezić, Ž., & Elitzur, M. 2002, ApJ, 570, L9

Nenkova, M., Sirocky, M. M., Ivezić, Ž., & Elitzur, M. 2008a, ApJ, 685, 147

Nenkova, M., Sirocky, M. M., Nikutta, R., Ivezić, Ž., & Elitzur, M. 2008b, ApJ, 685, 160

Noll, S., Mehlert, D., Appenzeller, I., et al. 2004, A&A, 418, 885

Papovich, C., Dickinson, M., Ferguson, H. C., et al. 2004, ApJ, 600, L111

Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2002, AJ, 124, 266

Peterson, B. M. 1997, An Introduction to Active Galactic Nuclei

Peterson, B. M., Ferrarese, L., Gilbert, K. M., et al. 2004, ApJ, 613, 682

Polletta, M., Tajer, M., Maraschi, L., et al. 2007, ApJ, 663, 81

Prevot, M. L., Lequeux, J., Prevot, L., Maurice, E., & Rocca-Volmerange, B. 1984, A&A, 132, 389

Quadri, R. F., & Williams, R. J. 2010, ApJ, 725, 794

Rangel, C., Nandra, K., Laird, E. S., & Orange, P. 2013, MNRAS, 428, 3089

Rees, M. J. 1984, ARA&A, 22, 471

Reynolds, C. S. 1997, MNRAS, 286, 513

Reynolds, C. S. 1999, in Astronomical Society of the Pacific Conference Series, Vol. 161, High Energy Processes in Accreting Black Holes, ed. J. Poutanen & R. Svensson, 178

Richards, G. T., Strauss, M. A., Fan, X., et al. 2006, AJ, 131, 2766

Rodighiero, G., Daddi, E., Baronchelli, I., et al. 2011, ApJ, 739, L40

Rosario, D. J., Mozena, M., Wuyts, S., et al. 2013, ApJ, 763, 59

Salimbeni, S., Castellano, M., Pentericci, L., et al. 2009, A&A, 501, 865

Salvato, M., Hasinger, G., Ilbert, O., et al. 2009, ApJ, 690, 1250

Salvato, M., Ilbert, O., Hasinger, G., et al. 2011, ApJ, 742, 61

Santini, P., Fontana, A., Grazian, A., et al. 2009, A&A, 504, 751

—. 2012, A&A, 538, A33

Santini, P., Ferguson, H. C., Fontana, A., et al. 2015, ApJ, 801, 97

Schmidt, M. 1968, ApJ, 151, 393

Schmidt, M., & Green, R. F. 1983, ApJ, 269, 352

Schödel, R., Ott, T., Genzel, R., et al. 2002, Nature, 419, 694

Scoville, N., Abraham, R. G., Aussel, H., et al. 2007, ApJS, 172, 38

Silk, J., & Rees, M. J. 1998, A&A, 331, L1

Silverman, J. D., Mainieri, V., Lehmer, B. D., et al. 2008a, ApJ, 675, 1025

Silverman, J. D., Green, P. J., Barkhouse, W. A., et al. 2008b, ApJ, 679, 118

Soltan, A. 1982, MNRAS, 200, 115

Spergel, D. N., Verde, L., Peiris, H. V., et al. 2003, ApJS, 148, 175

Steidel, C. C., Adelberger, K. L., Giavalisco, M., Dickinson, M., & Pettini, M. 1999, ApJ, 519, 1

Steidel, C. C., Adelberger, K. L., Shapley, A. E., et al. 2003, ApJ, 592, 728

Steidel, C. C., Giavalisco, M., Pettini, M., Dickinson, M., & Adelberger, K. L. 1996, ApJ, 462, L17

Sutherland, W., & Saunders, W. 1992, MNRAS, 259, 413

Tanaka, Y., Nandra, K., Fabian, A. C., et al. 1995, Nature, 375, 659

Tully, R. B., Pierce, M. J., Huang, J.-S., et al. 1998, AJ, 115, 2264

Ueda, Y., Akiyama, M., Ohta, K., & Miyaji, T. 2003, ApJ, 598, 886

Urry, C. M., & Padovani, P. 1995, PASP, 107, 803

Vanzella, E., Giavalisco, M., Dickinson, M., et al. 2009, ApJ, 695, 1163

Virani, S. N., Treister, E., Urry, C. M., & Gawiser, E. 2006, AJ, 131, 2373

Volonteri, M. 2010, A&A Rev., 18, 279

Williams, R. J., Quadri, R. F., Franx, M., van Dokkum, P., & Labbé, I. 2009, ApJ, 691, 1879

Wolf, C., Meisenheimer, K., Rix, H.-W., et al. 2003, A&A, 401, 73

Wuyts, S., Labbé, I., Schreiber, N. M. F., et al. 2008, ApJ, 682, 985

Wuyts, S., Labbé, I., Franx, M., et al. 2007, ApJ, 655, 51

Xue, Y. Q., Brandt, W. N., Luo, B., et al. 2010, ApJ, 720, 368

Xue, Y. Q., Luo, B., Brandt, W. N., et al. 2011, ApJS, 195, 10

Zheng, W., Mikles, V. J., Mainieri, V., et al. 2004, ApJS, 155, 73

Acknowledgements

I would like to thank all the people who made this thesis possible. First of all, I would like to express my biggest gratitude to my tremendous supervisor Dr. Mara Salvato who always has endless energy and ideas. I will always memorize the moment that the brilliant idea of the AGN/galaxy decomposition came up on the way you brought to the hospital. Thanks you Mara for teaching me with your unbelievable patience during my doctoral study. I am grateful to Prof. Dr. Paul Nandra who supported me and wisely guided me on the road of the scientific research. I would like to thank Dr. Antonis Georgakakis and Dr. Marcella Brusa for all the useful discussions we had. I really appreciate the research environment in our group and it is my honor to work with all these extraordinary scientists.

I would like to thank Dr. Sotiria Fotopoulou who patiently taught me using the *LePhare* from the very beginning, Dr. Johannes Buchner who helped me greatly on the X-ray-to-optical/IR associations, and Prof. Dr. Ralf Bender who provide me the valuable galaxy templates to build the AGN-galaxy hybrids. I would like to thank my lovely officemates Giulia Mantovani, Marie-Luise Menzel and super-kind friend Matteo Fossati. Thank you for sharing all the happiness and sadness during my entire bittersweet doctoral study. I also acknowledge the big support from Lai Yi and Wei Tien, who greatly help me in my most difficult time in Munich. Finally I am super grateful to my parents and family for their unlimited love, in particular thanks Yang for his huge support in my daily life in Germany.

The experience of studying and living in Germany makes me more open-mined. I have learned how to enjoy the life and love people around me. No matter where I go in the future, I will keep this precious memory in mind.